PROACTIVE BUSINESS INTELLIGENCE: DISCOVERING KEY PERFORMANCE INDICATORS WITH THE RULE EXTRACTION MATRIX METHOD

Robbin van den Houten and Marco Spruit
Department of Information and Computing Sciences
Utrecht University, The Netherlands

ABSTRACT

Key performance indicators have been around for a long time; organizations tend to provide their employees with metrics on strategic levels, such as market-share or profitability. By using metrics and performance indicators, organizations aim to influence the awareness of employees on targets. Performance increases are bound to occur, when employees are conscious of their personal results by metrics. However, an abundance of definitions of metrics overwhelm employees with values which are hard to grasp. By defining key performance indicators, relating to goals and strategic aspects of the organizations, meaningful indicators can be defined on an operational and tactical level, therefore, preventing cognitive overload.

This paper describes the extraction of key performance indicators by using the Rule Extraction Matrix (REM) Method. This method is constructed using several aspects of current data mining methods such as the CRISP-DM model, providing a solid base to determine the nominal value of a performance metric. The REM-method enables organizations to manage their performance by using key performance indicators extracted from data and literature. Thresholds to the key performance indicators are provided by business rules, to allow integration with systems and readability by analysts.

Validation of the REM-method is performed by a case study at a large Dutch Internet services provider and expert interviews were performed. The REM-method is an addition to the performance measurement field and further enhances an organization's ability to define their performance.

Keywords: Data mining, business intelligence, key performance indicators, business rules

INTRODUCTION

During times of recession, emphasis within organizations lies on available budgets and making the right decisions to guarantee the continuity of an organization. Pursuing improvement of financial results, production capacity or sales figures, results in several (ad-hoc) decisions. The results of these decisions have an increasing impact on the operational, tactical and strategic levels within an organization. While facing severe decisions which have an impact on the business, organizations need to obtain knowledge to undertake actions and ensure the effectiveness of decisions. Bierly, Kessler, & Christensen (2000) define knowledge as a clear understanding of information; transformation from data (raw facts) to information (meaningful, useful data) is specified as the process of gaining knowledge. In his view on a knowledge-based theory of the firm, Grant (1996) states that the critical input in production and primary source of value in organizations is knowledge. Currently, many chunks of valuable information, in the form of raw data and undefined performance indicators, remain unused within great pools of data. This paper describes the process of gaining knowledge from data available within databases to allow organizations to gain knowledge and use knowledge to support the decision making processes.

La Grouw (2009) suggests that many organizations have difficulties aligning (key) performance indicators, business rules and related decisions on a strategic, tactical and operational level. Key performance indicators are directly linked to a strategic outcome and regarded as key measures in terms of performance. Strategies are subject to change, in the event of a changing strategy, related key performance indicators must change accordingly to ensure tracking the effectiveness of the strategy. Furthermore, key performance indicators must have business rules attached to ensure the definition of performance thresholds, which are related to a key performance indicator. Walsh (1996) states that key performance indicators are decomposable in key performance drivers, plus key performance outcomes (key performance indicators that measure the progress towards corporate objectives). Organizations need to focus on key performance drivers, in order to influence key performance outcomes.

This paper describes the discovery of knowledge in the form of key performance indicators and business rules from historical data available in performance databases. Knowledge creation in the form of key performance indicators and associated business rules is acted upon by the analysis (data mining) of a (customer) performance database. The association of business rules draws on the same data as the mined key performance indicators. By mining the dataset and using the key performance indicators as input, variables for mining data result in associated business rules which constrain the key performance indicators. By providing organizations with key performance indicators and accompanying business rules, organizations are enabled to make the right decisions at the right moment and focus their attention.

This paper describes the extraction of relations from raw data between available key performance indicators and business rules, resulting in knowledge expansion. This paper's goal is to present the reader with a method by which organizations are able to extract key performance indicators and business rules from analyzed historical data, the method and its resulting output could serve as input for a (procedural) decision support system.

BACKGROUND

An overall umbrella definition for the two main topics within this research is business intelligence, well-known within most organizations due to the fact that success of an organization is dependent on the ability of an organization to make use of all actionable information. While more and more data is stored, organizations are challenged with the analysis of this ever-growing information store, combined with the fact that organizations are becoming more knowledge-centric. This leads to the access of a large number of employees to available knowledge within an organization, plus amounts to challenges of acting on information available within the firm (Cody, et al., 2002). The BI-FIT model is one notable attempt to address these issues (Tijssen, Spruit, Ridder & Raaij, 2010).

Research on business intelligence, as introduced in 1989 by Howard Dressner, describes concepts and methods to improve business decision making by using fact-based support (Negash & Grey, 2008). More recent definitions of business intelligence include the following "the leveraging of a variety of sources of data, as well as structured and unstructured information to provide decision makers with valuable information and knowledge"(Sabherwal & Becerra-Fernandez, 2010), "a data-driven decision support system that combines data gathering, data storage and knowledge management with analysis to provide input to the decision process" (Negash & Grey, 2008) and "computer-based techniques used in spotting, digging-out, and analyzing 'hard' business data, such as sales revenue by products or departments or associated costs and incomes" (Business Dictionary, 2009).

While some of the definitions for business intelligence complement each other, definitions mainly used by software vendors show contradictions such as regarding business intelligence as a technology instead of a method. As the amount of definitions and terminology grow larger and larger, resulting in less uniformity of this terminology and definitions, a boundary for this research needs to be provided. The boundary set on business intelligence applicable to this research is based on Figure 1: Data to wisdom traversal (Fugante, 2008) below. The (bounded) definition used in this research for business intelligence is the organizations' ability to construct knowledge from structured and unstructured data available in (legacy) systems.

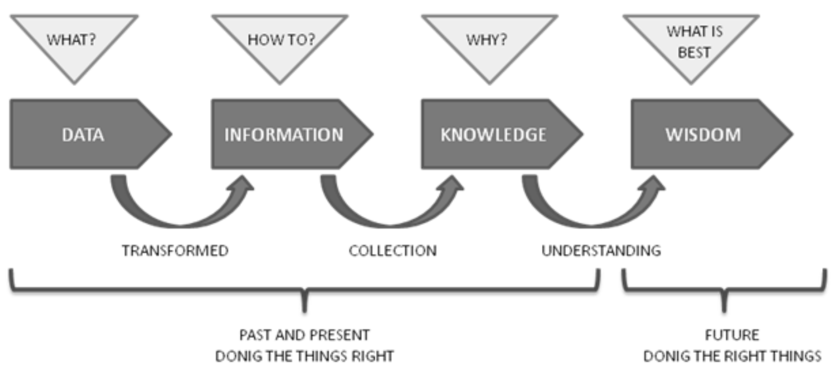

Figure 1. Data to wisdom traversal (Fugante, 2008).

Cody, Kreulen, Krishna & Spangler (2002) deem business intelligence and knowledge management as contributors to the improvement of quantitative and qualitative value of the knowledge available to decision makers. Combining business intelligence and knowledge management technologies enables the use of text mining and data searching algorithms, to extract information from textual data, and thus, attributing to the knowledge available to decision makers.

Extract, transform and load data – Knowledge management is an essential building block of this extraction, transformation and loading-process, since it is related to business intelligence. With knowledge and content management technologies used for searching, classifying and extracting information from data, knowledge management contributes to the data to information traversal. The technologies include clustering of information, taxonomy building, classification of data and summarization. The process of extracting, transforming and loading data typically entails the following (Aertsen, 2010) (Negash & Grey, 2008):

- Cycle initiation
- Build reference data
- Extract (from sources)
- Validate
- Transform (clean, apply business rules, check for data integrity, create aggregates or disaggregates)
- Stage (load into staging tables, if used)
- Audit reports (for example, on compliance with business rules. Also, in case of failure, helps to diagnose/repair)
- Publish (to target tables)
- Archive
- Clean up.

Next to business intelligence, emphasis for this research lies with the domain of data mining. In order to find patterns from the data that can attribute to the construction of key performance indicators and business rules, a subset of methods available within both domains is used. A formal definition for data mining applicable to this research is needed. Comparable to business intelligence, the term is used for a multiple of activities, such as the designation of any manual search of data, query assisted searching from a database management system, pattern visualization through human activity or the automated generation of transaction reports/any automated data correlation from these transaction reports (Fayyad, et al., 1996). The formal definition used within this research is the following: “Data mining is the process of searching and analyzing data in order to find implicit, but potentially useful, information. It involves selecting, exploring and modeling large amounts of data to uncover previously unknown patterns, and ultimately comprehensible information, from large databases” (Shaw, Subramaniam, Tan, & Welge, 2001) also referred to as knowledge discovery or data-driven discovery through knowledge verification or prediction/description (Fayyad, et al., 1996).

Data mining relies on the availability of large resources of data available through computational methods which are based on statistical analysis, decision trees, neural networks, rule induction, refinement and graphic visualization. While certain limitations arise when dealing with datasets reaching into multiple terabytes of data, data mining is also carried out using representative samples of data.

Traditionally, data analysis occurs by one or more analysts becoming overly and intimately familiar with the data and act as an intermediary between the (raw) data and the (end-) user. The current data mining solutions are coined towards solving one of this information capturing and storing era's problems knowing information overload. Huge capacities of data storage, for example the World Wide Web, can be viewed as huge low-level data volumes. These volumes of data are typically difficult to understand and often digested into information in the form of reports or predictive models. Data mining methods are applied to these pools of data storage to extract information and patterns and effectively attribute to knowledge, through prediction and description, available to the public or organizations by automatically producing useful information from large masses of (raw) data (Fayyad, et al., 1996) (Cooley, et al., 1997) (Pachidi, Spruit & Weerd, 2014).

CRISP-DM – An addition to knowledge discovery from databases is the cross industry standard process for data mining (CRISP-DM), which originated from collaboration between three major industrial players within the domain of data mining: DaimlerChrysler, NCR and SPSS. The trigger event for the development of this model is the lack of description on how to implement data mining results from an organizations' point of view.

The introduction of this model sets out its base not only on the development of an understanding of the application domain, but also the understanding of the business itself, next to that it adds steps to provide an organization with a deployment phase (Chapman, et al., 2000).

RESEARCH QUESTION AND APPROACH

In order to define the research question and scope for this research, the research goals have to be addressed; this section describes the research goals and the emerging research question.

Research Goals and Research Question

The research objectives entail the following:

- Propose a formative definition for key performance indicators to identify these among performance indicators, available in the field the organization is operating

- Establish a generalizable method, by which key performance indicators and business rules can be associated and aligned
- Identify current key performance indicators and associate these with available business rules, following the proposed method to allow for decision support and usage in the systems at hand
- Assess the method, by using the proposed case study analysis and explorative interviews.

The research addresses the above research goals by posing the following research question:

> "How Can Historical Data About Employee and Organizational Performance Be Transformed Into Business Rules That Define the Effects of Fluctuations In Performance Indicators?"

Approach

Since little research has been performed on combining key performance indicators with business rules, the development of a method is comparable to a wicked problem as described by Hevner, March, Park and Ram (2004), a wicked problem is characterized by the following constraints:

- Unstable requirements and constraints based on ill-defined environmental contexts
- Complex interactions among subcomponents of the problem and its solution
- Inherent flexibility to change design processes, as well as design artifacts (i.e., malleable processes and artifacts)
- A critical dependence on human cognitive abilities (e.g., creativity) to produce effective solutions
- A critical dependence on human social abilities (e.g., teamwork) to produce effective solutions.

Several of these descriptions are applicable to the development of the proposed method for associating business rules and key performance indicators, knowing that existing models are available for performance measurement; however, these do not serve the purpose of incorporating business rules.

The steps in developing a method will take the approach of design science as shown in figure 2 taken from Takeda, Tomiyama, Yoshikawa, & Veerkamp (1990). The first process step in the design cycle is the awareness of the problem:

> "How can knowledge be derived from raw historical data available within an organization and support future decisions?"

The second step in the design cycle shows the hypothetical suggestion or answer to the problem. The third step portrays the creation of an artifact, based on the knowledge gained from the completion of the second step. The fourth step within the design cycle shows the evaluation phase, in which the artifact is evaluated and eventually elaborated on by the fifth step, following a normal scientific research approach.

An exploratory case study will be used as the strategy for qualitative analysis for this research. The goal for this case study is to extract from data which organizational performance metrics are influenced by individual performance and how this influence can be measured and presented in the form of business rules. Yin (2003) defines the case study as an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident. Another definition of the case study is provided by Flyvbjerg (2006), this author states that a case study is an intensive analysis of an individual unit (e.g., a person, group, or event) stressing developmental factors in relation to context.

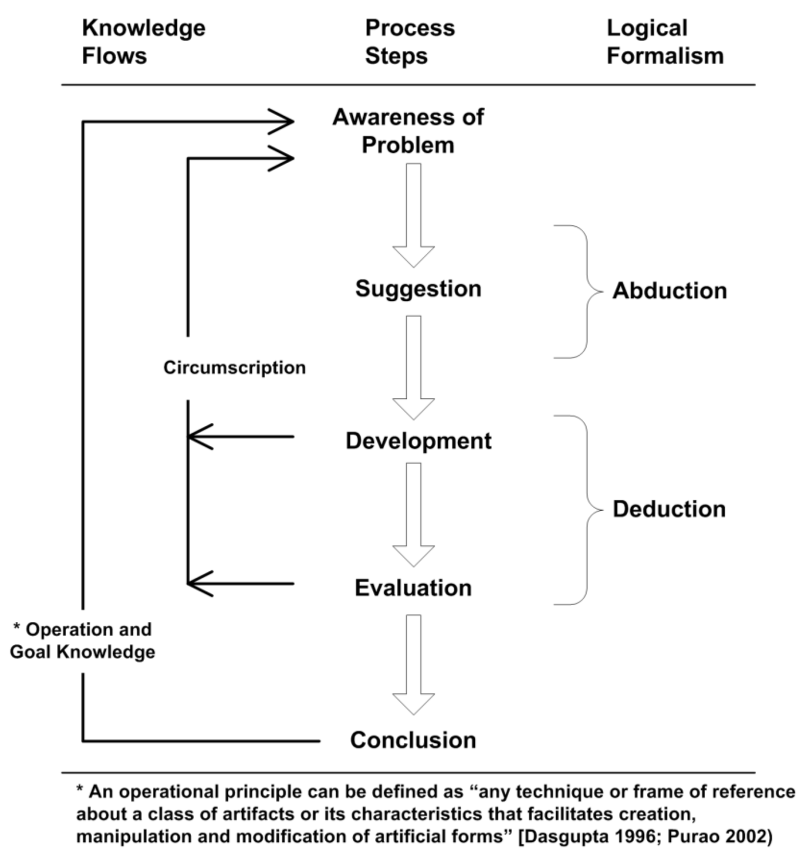

Figure 2. Reasoning in the design cycle (Takeda, et al., 1990).

Within this case study the unit of analysis is the organizational cooperation of a customer service center with the telecommunications provider. The data for this case study is collected by using two separated sources, the first being a database containing data concerning the organization with regard to project performance, the data collection in the first database is fully automated. There is no human interaction other than retrieving the datasets. The second is a database filled by quality employees who monitor service requests on a project using a five-point Likert scale and rank a monitored service request based upon thirteen quality aspects. A random selection of service requests that are available for ranking is presented to the quality employee. One of the main criteria for the amount of service requests rated is the ranking should include at least three service requests per employee per month. In order to increase the accuracy of the results, the employees collecting the data are using an interface detailing the elements a quality aspect should contain in order to receive a specific score on the five point scale.

From applying data mining algorithms one should be able to find independent variables influencing dependent variables, by pattern matching independent variables and concluding whether their change affects the dependent variables. The findings are interpreted by statistical tests, as well as proposing the data to respondents and determining if the relations that have been derived from the data explain reality. In other words, the explanations for the effects suffice, or the effects are subject to rival explanations. The case study will show the data collected from various sources is connected by using the method. The main theory for this case study is that by using the method changes in employee performance and their influences on organizational performance become visible. These influences will be translated in performance indicators and business rules.

Project objectives include selecting a sample that represents reality; this means the data from the second source should be cleansed from any anomalies and should still be a sample large enough. The average monthly size of the population size (service requests) is 24000, averaging 600 (200 employees * 3) service requests scores monthly. This represents a sample large enough to represent the reality based upon a 95% confidence level. Service requests that are not ranked completely will not be included in the analysis of the data.

The first relevant test for a case study is construct validity, essentially this entails establishing correct operational measures for the concepts being studied. Within the case study carried out in the research, construct validity is met by establishing multiple sources of evidence, the hypothesis is tested using data from the field generated by automated operational systems as well as employees trained in scoring service requests (in this case data collection). A chain of evidence is maintained with regard to data sources as well as data location(s).

The second test concerns internal validity. Yin (2003) states internal validity is only a concern when performing causal or explanatory case studies. This is the case in case studies where an investigator tries to determine if event x led to event y. The research also looks at event x leading to event y (whether changes in employee performance lead to changes in organizational performance); however, this is only the case for the supporting data. The research question and applicable theory try to answer how we can derive those influences from historical data and thus the form of the case study is deemed as explanatory.

The third test in doing case study research concerns the external validity; Yin (2003) explains external validity as establishing the domain to which a finding can be generalized. With regard to the research question the tested theory (the method is suitable to extract business rules and performance indicators from historical data) can be generalized as applicable to customer service centers or comparable organizations where employee and organizational output are closely monitored.

The selected case is representative among many typical comparable projects. In essential the mode of generalization concerns the analytic generalization; in this case a previously defined theory is used as the template, by which the empirical results of the case study are compared.

The fourth test applicable to the quality of empirical research is reliability; reliability is tested by demonstrating the operations of study are repeatable with the same results. To achieve reliability two techniques are available, constructing a case study protocol and the creation of a case study database, the technique demonstrated within the previous sections concerns the case study protocol. By providing the user with a clear definition of the case study protocol we aim to meet the fourth test.

Case study data will be presented using templates provided by the data mining tools. Key performance indicators and business rules derived from data are visualized in order to present them to respondents which in turn provide input for expert interviews.

The REM-Method

This chapter addresses the method used in this research. Figure 3 depicts the Rule Extraction Matrix Method, the figure shows the steps to extract information in the form of key performance indicators and business rules associated through the key performance matrix. The image displayed below is the starting point for the method. Based on literature study and the current standards in data mining known as CRISP-DM (KDnuggets, 2007), the method describes the steps, following the approach of Brinkkemper (1996), that need to be followed to empower organizations to extract information in the form of key performance indicators and business rules. Note that since all process steps will be performed on-site, we favor the CRISP-DM method in this case over the Three Phases Method (Vleugel, Spruit & Daal, 2010). A more detailed description of the subparts, found by the patterned lines covering an area (example given: business understanding), is displayed in the sections that follow this introduction.

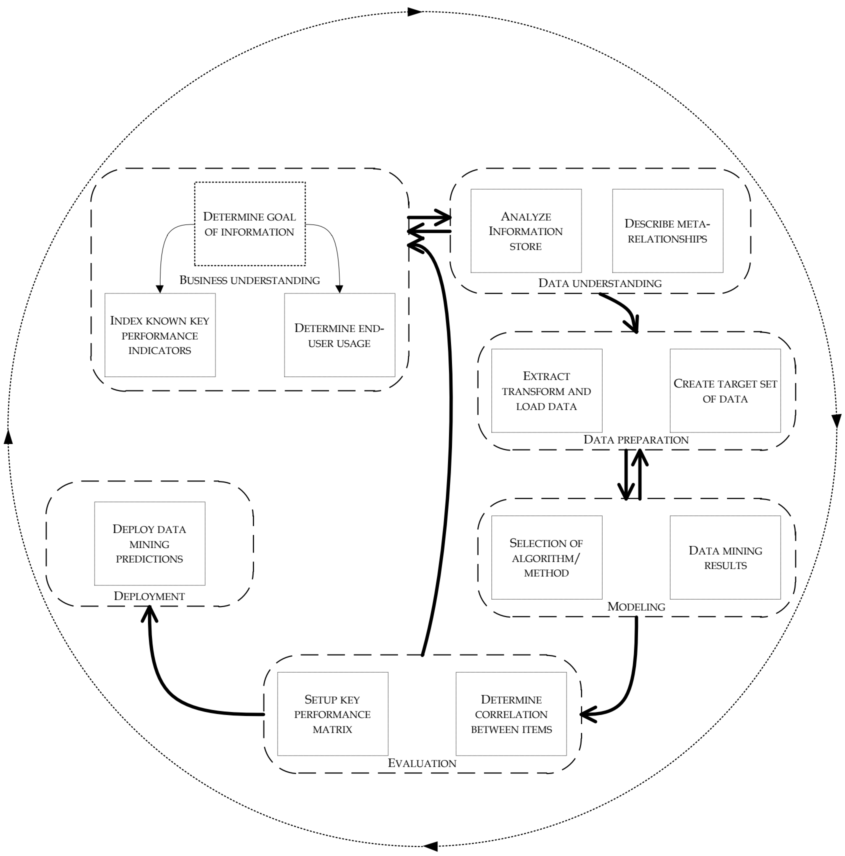

Figure 3. Rule Extraction Matrix (REM) Method.

The table below displays the phases taking place within the REM-method; it consists of the steps which relate to the phases as listed in figure 3, the main preferable actors of these steps and the results following the execution of each process step.

Business Understanding

While the REM-method is a continuous cycle, many organizations will find the first steps towards the extraction of key performance indicators and deployment of business rules begin with the business understanding. Chapman et al. (2000) describe business understanding as the initial phase concentrating on project objectives and requirements from a business perspective. To integrate the extraction of business elements (key performance indicators and business rules) the method is extended by determining the end-user usage by means of the MAD-framework (Eckerson, 2009)

Table 1. REM Tasks

Phase	Step(s)	Actor	Result(s)
Business understanding	Determine goal of information	Analyst	Information analysis containing goals, information on business mission/case.
	Index known key performance indicators	Analyst/Management	List of existing business rules and key performance indicators containing the main indicators for the business/field.
	Determine end-user usage.	Analyst/Management	Target-groups, how will the results in the form of business rules be used by the business.
Data understanding	Analyze information store	Data Analyst	Overview of data information sources, trigger hypothesis.
	Describe meta-relationships	Data Analyst	Provide an overview of the meta-relationships between the dataset(s).
Data preparation	Extract, transform and load data	Data Analyst	Clean draft dataset which is filtered, extracted and enriched.
	Create target set of data	Data Analyst	Errorless target dataset.
Modeling	Selection of algorithm/data mining method	Data Analyst	Applicable data mining algorithm based on available target dataset.
	Data mining results	Data Analyst	Results following the selection and appliance of the data mining algorithm.
Evaluation	Setup key performance matrix	Data Analyst	Draft key performance matrix containing an overview of performance indicators.
	Determine correlation between items	Data Analyst	Final key performance matrix containing correlated items through mining and research.
Deployment	Deploy data mining predictions	Data Analyst	Translated results towards business rules (which items influence which).

By finding an answer on the use of the results, essentially answering the question: “will they be incorporated in reports or an automated system?” The above, supported by a list of current key performance indicators, results in the following steps to realize the business understanding, detailed in the figure below.

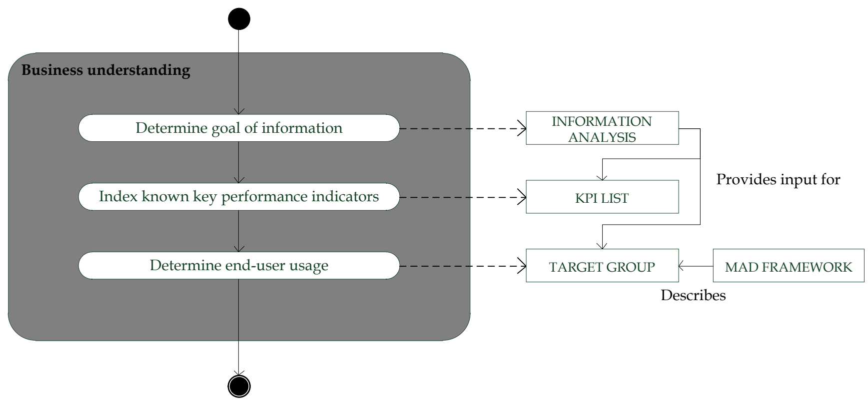

Figure 4. Process deliverable diagram: Business understanding.

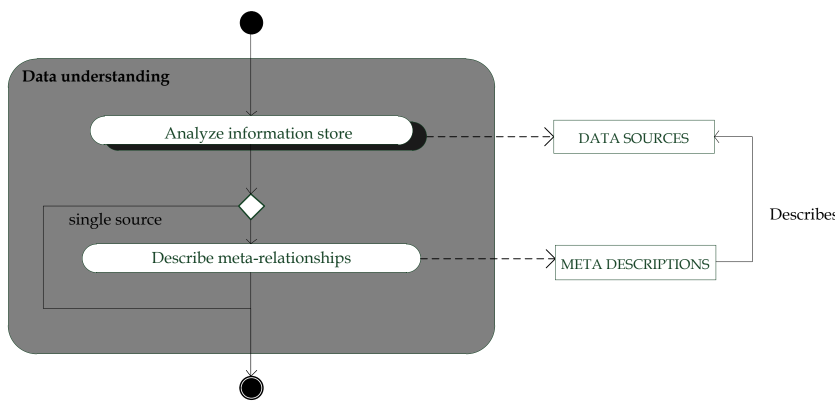

Figure 5. Process deliverable diagram: Data understanding.

Data Understanding

Next in the cycle is the data understanding. Chapman et al. (2000) describe the data understanding phase as essential towards the initial data collection and identifying quality problems, in order to perform this, the preceding business understanding step will need to be performed since the data analyzed is heavily influenced by the goals and key performance indicators that have been found in the previous step.

Data Preparation

The following step in the cycle is the data preparation phase. Chapman et al. (2000) determined the data preparation phase include tabulating the data, selecting attributes and cleaning the data of errors/malformed entries. This phase is heavily dependent on the data available in the information store, and as such it is burdened with the tasks of extracting, transforming and loading data, as well as creating a target set of data applicable for analysis.

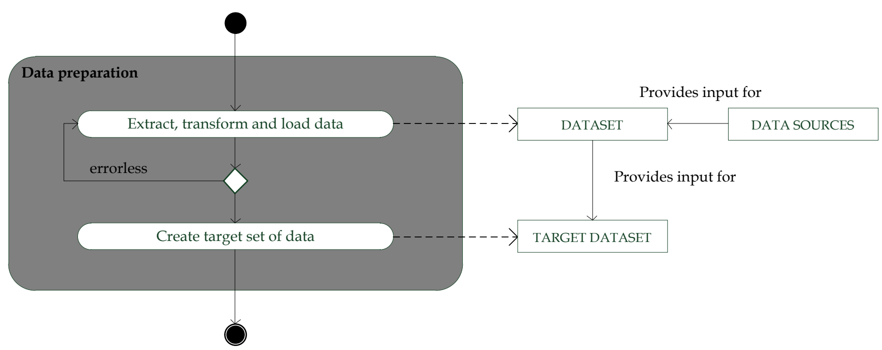

Figure 6. Process deliverable diagram: Data preparation.

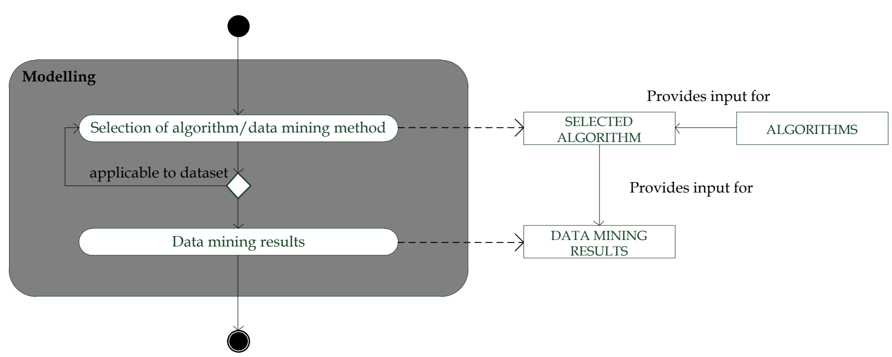

Figure 7. Process deliverable diagram: Modeling phase.

Modeling Phase

The step following the data preparation is the actual modeling phase. Chapman et al. (2000) describe the modeling phase as the selection of various modeling techniques and parameter calibration. The modeling phase consists of the selection of an algorithm/method that is applicable for the data at hand and the data mining results following the selection of the algorithm/method.

Evaluation Phase

Following the modeling phase, the evaluation phase reviews the results from the data mining step. Chapman et al. (2000) deem the phase to be responsible for reviewing the results and analyzing the success or failure of the constructed model. Within this phase the construction of a key performance matrix and the evaluation of item correlation take place.

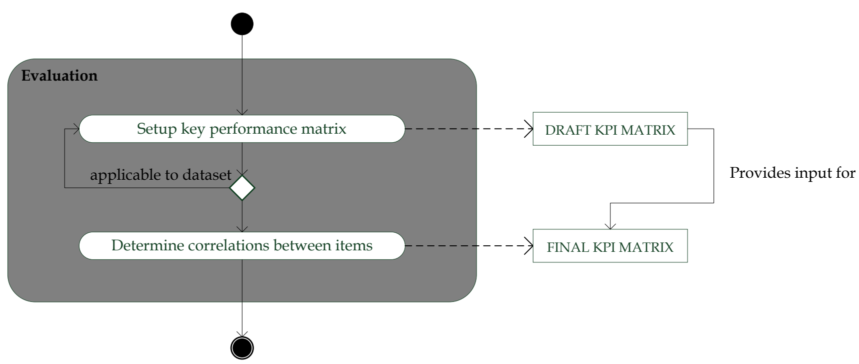

Figure 8. Process deliverable diagram: Evaluation phase.

Deployment Phase

The final step in the cycle is the deployment phase; however, since the steps take place in a cycle this does not mean that the project has come to an end, often calibration/updates take place which entail a new run through the cycle to optimize results. The deployment phase entails the incorporation of data mining predictions, in this research this means displaying of information towards the parties involved (determined within the business understanding).

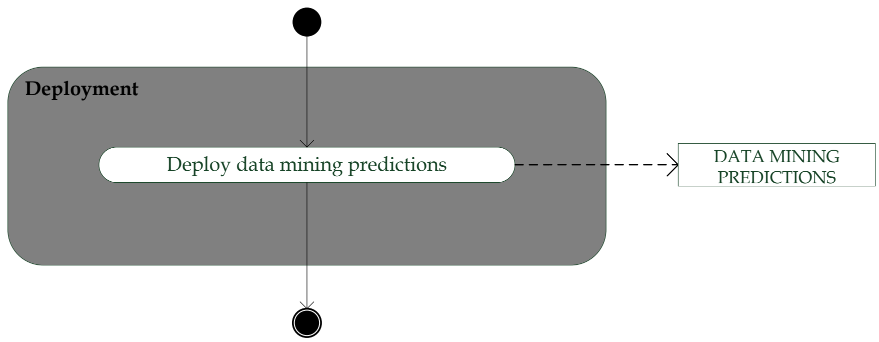

Figure 9. Process deliverable diagram: Deployment phase.

CASE STUDY

The following section outlines the case study carried out to determine if the REM-method suffices in determining the relations between key performance indicators and displaying these relations using business rules. The introduction shortly describes the organizations' structure, as well as the case study set-up including the data sources used for the data mining part of the case study. The sections following the introduction of the case study will follow the phases as defined in the REM-method.

Introduction

The case study is performed at a customer service center and analyzes the cooperation with a large Dutch telecommunications provider. The organization consists of approximately 1000 employees (750 F.T.E) acting worldwide at market-segments such as healthcare, telecommunications and financial services. Current analysis on key performance indicators is one of the pitfalls in managing the processes within the organization. Analysis is often performed by static excel reports that do not represent long-term information but cover ad-hoc situations and questions that are asked by the service provider's outsourcing managers. To stay ahead of problems and have an answer ready to long-term questions and situations the organization turns towards analyzing its current information sources. Summarizing the case leads to the following: customers need information on successful factors for their customer's inquiry to be handled successfully. By analyzing key performance indicators from a business perspective, determining their measurement by using business rules we aim to provide the information needed.

Business Understanding

The first step in the REM-method is the Business Understanding phase. From this phase we derived the current key performance indicators that are used in the field the organization is operating in. Current key performance indicators resulted in a long list of (dependent) variables, from which was concluded that not all performance indicators in the field were measured due to little business relevance. The business understanding phase also consists of determining the end-usage for the proposed results of the method. From this we determined that the proposed end-users were automated systems and project managers, thus the form of business rules needs to be addressed towards these uses.

Data Understanding

From the business understanding phase we derived a long list of dependent variables, given the data available from the systems, this resulted in a short list of variables available for analysis. The dataset combines three available data-stores resulting in a data-structure containing the variables measured, brought back to the weekly measuring level, the smallest level one of the parties could report.

The final dataset is constructed from 451 days of quality monitoring entries amounting into 3971 unique records. However, this is still a small (project) subset of the total amount of service requests, which are performed for the service provider since quality monitoring per call is not always possible due to time limits, samples (or quality monitoring entries) are providing a large enough sample as detailed in section 3.2.

By analyzing the previously mentioned datasets and keeping the research question in mind, the following trigger hypothesis is defined:

> An increase in independent variables measured at the employee level results in a positive or negative effect on dependent variables measured at the organizational level.

The above allows combining systems and tries to find specific relations between performance measured by the provider and the service center.

Data Preparation

The third step of the REM-method covers the data preparation phase, this phase takes care of the dataset, the phase mainly results in a clean errorless dataset on which (several) data mining algorithms can be executed.

The process of extraction, transforming and loading of data is based on the research by Aertsen (2010) and Negash & Grey (2008) and includes the following steps:

- Address unknown or NULL values
- Remove duplicates
- Remove malformed data entries
- Remove incomplete entries
- Applying (existing) business rules
- Export datasets

The above steps are performed by tools for data cleansing, plus exporting the datasets, using settings to prevent duplicates and malformed data entries.

The cleansed exported datasets derived in the previous step, in the data preparation phase, are then transformed to allow analysis by the selected data mining tool. For this case study Rapid Miner 5.1 (Open-source, AGPL) is used, allowing for several steps in the data mining process and the offering of data manipulation by extraction, transforming and loading of data. The errorless dataset that results from the extraction, transaction and loading of data is fed to the data mining tool. By algorithmic analysis, errors are analyzed and the data mining tool is provided with the right fields and meta-information for those fields.

Modeling Phase

Focus of algorithmic analysis lies within which values are influenced by related key values and how strong the effects are. The influences and effects are measured by creating decision trees and business rules from the variables at hand. In order to create an overview of the dataset and to see which variables suffice for analysis descriptive methods are used that allow summarization of the datasets properties. Through analysis it is learned that several variables are numerical; however, decision trees need categorical data to allow data to be analyzed and structured in a decision tree. In short, translation is needed from numerical to categorical. In order to overcome this problem, another univariate descriptive method is needed to allow for binning of the numerical variables.

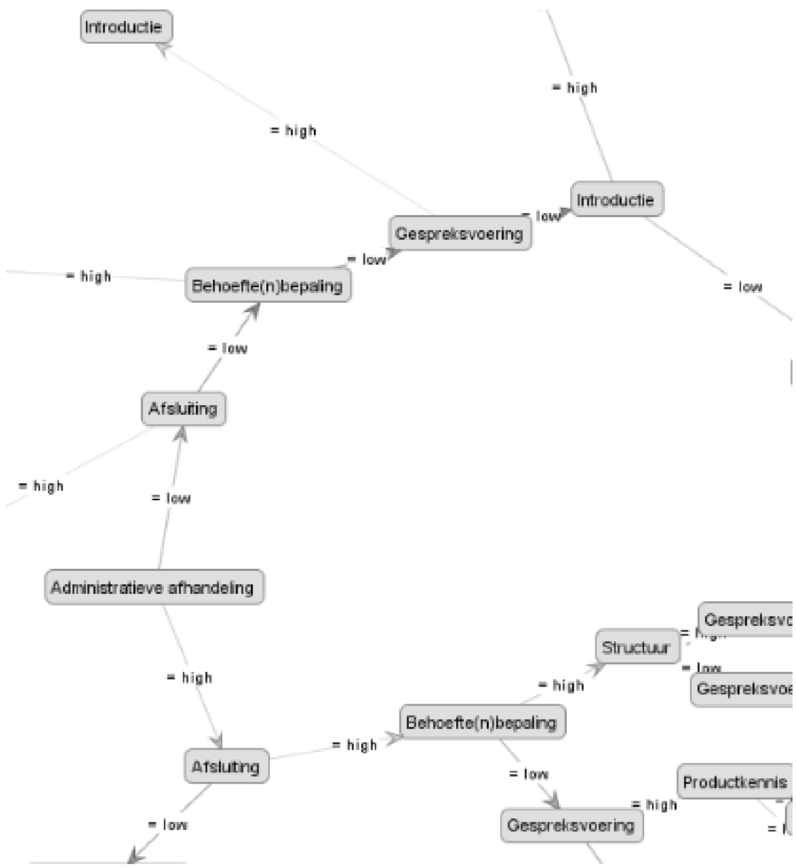

Figure 10. Example result modeling phase.

Using the results of the binning process, the data is relayed to the predictive methods used in the data mining tool. From the dataset a decision tree is created that predicts the target variable (key performance indicator) from the variables fed to the method, the target variable in this case is one of the previously derived key performance indicators from literature. The creation of the decision tree is performed through classification of variables, fed to the method and predicting the target variable through the before mentioned classification. Finally, to allow for feeding the results to the system and make them readable by analysts, business rules need to be created from the decision trees. The translation from decision trees towards business rules is performed by feeding the data to association rules methods that allow finding patterns. Example of a resulting business rule:

- if Behoefte(n)bepaling = high and Leiding behouden = high and Luisteren = high then low
- if Behoefte(n)bepaling = high and Leiding behouden = high and Luisteren = low then high
- if Behoefte(n)bepaling = high and Leiding behouden = low then high
- if Behoefte(n)bepaling = low then low

Evaluation Phase

The key performance matrix shows influences that have been derived from the various business rules and decision trees. The creation of the final key performance matrix is dependent on the correlations between each item. Using this matrix, the customer service provider will be able to direct its focus by using a combination of the effects measured above. The following figure shows the correlations between dependent variables and independent variables:

Correlations

	Service-level	Call time	Conversie
Behoeftenbepaling	,630**	,610**	,594**
Gespreksvoering	,627**	,575**	,241**
Klanttevredenheid	,122**	,352**	-,092**
Leiding behouden	-,076**	-,380**	,075**
Luisteren	-,126**	-,251**	-,142**
Productkennis	,202**	,308**	-,116**
Proposities en Analyse	,507**	-,475**	,521**

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 11. Correlation dependent and independent variables.

From the above figure one can derive which variables have the largest influence on critical dependent variables. An example of this is call time, which in turn influences other dependent variables. Call time is positively influenced by “behoeftebepaling” and “gespreksvoering” and negatively influenced by “proposities and analyse”. When aiming for a low call time (decrease is desired), the above would entail that “behoeftebepaling” and “gespreksvoering” should be low and “proposities en analyse” should be high.

Table 2. Final key performance matrix

Variable	Dependent variable	Effect
Behoeftebepaling	Service-level	Positive (Increase)
Gespreksvoering	Service-level	Positive (Increase)
Propositie en Analyse	Service-level	Positive (Increase)
Behoeftebepaling	Call-time	Negative (Increase)
Gespreksvoering	Call-time	Negative (Increase)
Propositie en Analyse	Call-time	Positive (Decrease)
Behoeftebepaling	Conversie (totaal)	Positive (Increase)
Propositie en Analyse	Conversie (totaal)	Positive (Increase)

Deployment Phase

The data mining predictions, allow integration with systems used to monitor performance at an individual level. Using these rules allows integration with systems used that help with SWOT-analysis, essentially applying the analyze-functionality. Next to the implications towards systems these results have, the key performance matrix, offers project managers an operational view of which performance indicators influence which key performance indicators. This allows for a better analysis and advice towards the customer, not based on guesswork but by using historical facts to underline relations between performance variables.

CONCLUSION

By combining performance on an employee level and evaluation of performance on an organizational level, performance indicators can be found.

These in turn lead to key performance indicators based on organizational criteria. The fluctuations in key performance indicators can then be used to calculate the influences they exert on other key performance indicators on an organizational scale. By the use of the REM-method and the consequent phases embodied in the method the above can be applied to a variety of (historical) data. This research found influences in the historical data and by the use of the REM-method translated these influences into business rules, which in turn can be used to further drive the performance of the various projects within the organization.

By the application of the REM-method, new business rules were extracted from the data at hand. The extensions applied to the CRISP-DM model allowed for successful application to the case study. The extensions fitted to the model allow obtaining a better understanding of the data and the processes that accompany the processing of data and information extraction. The method's main deliverable is the performance matrix which shows the correlations between variables and the business rules supporting the extraction of these correlations. When looking at the performance matrix and the responses of the experts in the field, it can be said that these results closely adhere to the real world scenarios.

Given the responses of experts during the interviews, the performance matrix is a welcome candidate to evaluate performance with and by the customer. It can be used to describe relations between variables, as well as by applying the right dashboard techniques, display historical relations and their development. The main focus lies with the relations defined between dependent and independent variables and the resulting business rules. These were interpreted by the respondents and deemed true, even if assumptions made in the past were telling them otherwise. Summing up the pro's and con's of the performance matrix, it can be said that the research goals were met.

All in all, the research achieved what it set out to do, through the execution of a case study a set of key performance indicators are defined and these can be used to evaluate and further drive performance on an employee level, as well as an organizational scale. The case study leans heavily on the results of one project/campaign, due to time constraints. It would be better to combine the results of multiple campaigns to create a model that represents the whole organization, instead of just one campaign.

REFERENCES

Aertsen, J. (2010, 02 18). *Extract transform load*. Consulté le 05 12, 2010, sur Business intelligence compendium: http://www.business intelligencecompendium.com/index.php?title=Extract,_transform,_load

Bierly, P. E., Kessler, H. E., & Christensen, E. W. (2000). Organizational learning, knowledge and wisdom. *Journal of Organizational Change Management , 13* (6), 595-618.

Brinkkemper, S. (1996). Method engineering: engineering of information systems development methods and tools. *Inf. and Software Technology , 38* (4), 275–280.

Business Dictionary. (2009). *Business intelligence (BI)*. Consulté le 04 10, 2010, sur Business Dictionary: http://www.businessdictionary.com/definition/business-intelligence-BI.html

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., et al. (2000). *CRISP-DM 1.0 Step-by-step data mining guide.* Consulté le 05 20, 2010, sur CRISP-DM Consortium: http://www.crisp-dm.org/CRISPWP-0800.pdf

Cody, W. F., Kreulen, J. T., Krishna, V., & Spangler, W. S. (2002). The integration of business intelligence and knowledge management. *IBM Systems Journal , 41* (4), 697-713.

Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the World Wide Web. *Internatational Conference on Tools with Artificial Intelligence*, (pp. 558-567). Newport Beach, CA.

Eckerson, W. W. (2009). *Beyond Reporting. Delivering Insights with Next-Generation Analytics.* The Data Warehousing Institute.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *Artificial Intelligence Magazine* .

Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research . *Qualitative Inquiry , 12* (2), 219-245.

Fugante, V. (2008, 09 6). *From Data to Wisdom*. Consulté le 04 28, 2010, sur http://blogs.southworks.net/vfugante/2008/09/06/from-data-to-wisdom/

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal , 17*, 109-122.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly , 28* (1), 75-105.

KDnuggets. (2007). *Poll: Data Mining Methodology*. Consulté le 05 23, 2010, sur http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm

La Grouw, G. (2009, 09 15). *Aligning KPI, Business Rules and Decisions*. Consulté le 10 11, 2009, sur Logical Organization Blog: http://thelogicalorganization.com/blog/2009/09/aligning-kpi-business-rules-and-decisions/

Negash, S., & Grey, P. (2008). *Handbook on Decision Support Systems 2.* Heidelberg: Springer Berlin.

Pachidi, S., Spruit, M., & Weerd, I. van der (2014). Understanding Users' Behavior with Software Operation Data Mining. *Computers in Human Behavior, 30*, 583–594.

Sabherwal, R., & Becerra-Fernandez, I. (2010). *Business Intelligence.* Westford: Blackwell Publishers.

Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems , 31*, 127-137.

Takeda, H., Tomiyama, T., Yoshikawa, H., & Veerkamp, P. J. (1990). Modeling design processes. *AI Magazine*, 37-48.

Tijssen, R., Spruit, M., Ridder, M. van de, & Raaij, B. van (2011). *BI-FIT: Aligning Business Intelligence end-users, tasks and technologies*. In : Enterprise Information Systems Design, Implementation and Management: Organizational Applications, 162-177.

Vleugel, A., Spruit, M., & Daal, A. (2010). Historical data analysis through data mining from an outsourcing perspective: the three-phases method. *International Journal of Business Intelligence Research, 1*(3), 42-65.

Walsh, P. (1996). Finding key performance drivers: Some new tools. *Total Quality Management* , 509-519.

Yin, R. K. (2003). *Case Study Research: Design and Methods.* Sage Publications.