MASTER THESIS DOCUMENT

PROACTIVE BUSINESS INTELLIGENCE

DISCOVERING KEY PERFORMANCE INDICATORS AND ASSOCIATED BUSINESS RULES FROM HISTORICAL DATA USING DATA MINING TECHNIQUES

AUTHOR:	ROBBIN VAN DEN HOUTEN
STUDENTNUMBER:	3144208
FACULTY:	FACULTY OF SCIENCE
DEPARTMENT:	INFORMATION AND COMPUTING SCIENCES
SUPERVISORS:	DR. MARCO SPRUIT (UNIVERSITY UTRECHT)
	DR. SLINGER JANSEN (UNIVERSITY UTRECHT)

PREFACE

In front of the reader is the thesis on "*Discovering key performance indicators and associated business rules from historical data using data mining techniques*". The thesis was conducted at VANAD B.V., due to an unwelcomed accident, delays were inevitable.

I would like to use this opportunity to thank the people assisting me during the period preceding the start of the thesis and to those who supported me throughout the process of researching, writing the documents and relating tasks of this master's research. Firstly, I would like to thank my supervisor Dr. Marco R. Spruit for sending quick responses, providing advice and directing this thesis. Secondly, I would like to thank Joost Plugge, Kelly Ufkes and Arnoud Munneke from VANAD B.V. for providing me the possibility to finish my master at this organization, as well their endless efforts in stimulating me to graduate. In addition, I would like to thank the specialists for their responses and the people who provided the required data for this research.

Finally, I would like to thank my friends and family for helping me through hard times preceding the writing of this thesis and their efforts in stimulating me to finalize my work.

Robbin van den Houten

Utrecht,

Tuesday, July 17, 2012

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. Next to the key performance indicators that serve the strategic vision of an organization, tactical and operational levels are also influenced by metrics. By using metrics and performance indicators, organizations aim to influence the awareness of employees on targets on a personal and organizational level. Performance increases are bound to occur, when employees are made conscious of their personal results by using metrics. However, an abundance of current definitions of metrics and performance indicators 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, prevent cognitive and operational overload.

This research focuses on the key aspects of defining key performance indicators by studying literature containing metrics, the metrics studied are often used in the field in which an organization operates. These metrics are compared to the ones, which are currently employed within the organization's structure. As a result of comparing data and active metrics, a selection of key metrics is provided to the organization. The availability of these key metrics, enable organizations to focus on the business, instead of defining metrics.

The key performance metrics are extracted from literature and available data is used in the Rule Extraction Matrix (REM) Method. This method is constructed using several aspects of current data mining methods. One of the applicable data mining methods is the CRISPDM model, which provides a solid base to determine the nominal value of a performance metric. The method provides ways to bond the performance metrics, using business rules, and thus leaving room for an interpretation of these rules using a decision support system.

The REM method enables organizations to manage their performance by using bounded metrics and extract business rules, business rules are responsible for providing thresholds to the extracted key performance indicators. Validation of the REM method is performed by an extensive case study at a large Dutch Telephone and 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. Moreover, the REM method is applicable to a variety of organizations given the following requisite: historical performance measurements must be available.

Key words: key performance indicators, business rules, Rule Extraction Matrix method, REM, CRISPDM, performance measurement, decision support, key metrics, data mining, method fragments, method engineering.

TABLE OF CONTENTS

TABLE OF FIGURES

1. INTRODUCTION

During these times of recession, emphasis within organizations lies on available budgets and making the right decisions to guarantee the continuity of an organization. The pursuit of improving financial results, as well as production capacity or sales figures, results in several (ad-hoc) decisions, impacting the operational, tactical and strategic levels. While facing these severe decisions which have an impact on the business, organizations feel the need for obtaining the right amount of knowledge to undertake actions and ensure the effectiveness of a decision. Moreover, organizations need to define the impact of such a decision. 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. Usage of gained or extracted knowledge to establish and achieve goals, set by an organization, is described as wisdom. Information within an organization is often available in the form of (performance) metrics, some of these metrics might be directly available or hidden within databases; these metrics have the ability to define performance, in relation to previous obtained results, but also in relation to other organizations.

La Grouw (2009) notes 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. Since strategies are often subject to change, related key performance indicators must change accordingly to ensure tracking of 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) denotes that key performance indicators are decomposable in key performance drivers, plus key performance outcomes (key performance indicators that measure the progress towards corporate objectives) are both influenced throughout business processes. Organizations need to focus on key performance drivers, in order to influence key performance outcomes. As described in the first paragraph, construction of metrics in the form of (key) performance indicators and associated business rules amounts to knowledge creation within an organization. By defining key performance indicators and business rules organizations are enabled to make the right decisions at the right moment.

A widely quoted definition of a business rule is the following: ”a statement that defines or constrains some aspect of the business. It is intended to assert business structure or to control or influence the behavior of the business”. Business rules constrain what must or must not be the case, at any given moment it should be possible to determine, if the condition implied by the business rule in the form of a constraint is true in a logical sense, if not action should be taken. From a software perspective the constraint can be regarded as a boolean, regarded that the outcome of a constraint is either true or false, in turn sharing its characteristics with the boolean (Morgan, 2002).

This research proposes the discovery of knowledge in the form of key performance indicators and business rules from historical data available in performance databases. One example of knowledge creation in the form of a key performance indicator and associated business rule(s) could be the analysis (mining) of a customer database. Hypothetically, a customer database contains customer data (i.e. name, address, and current subscription(s)) historical interaction data (i.e. outcome of (previous) service requests, propositions and accepted/declined propositions). By data analysis of the database, an example of a key performance indicator could be 'accepted propositions', the associated business rule could be 'customers are willingly to accept a proposition when it has not been mentioned more than three times in previous interactions'. Figure 1 below shows the schematic approach of knowledge creation from historical data which is available in performance databases.

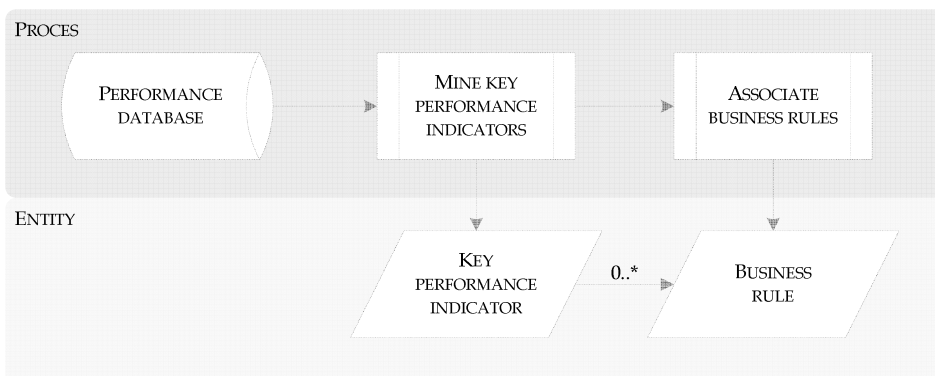

FIGURE 1: SCHEMATIC APPROACH KNOWLEDGE CREATION FROM HISTORICAL DATA

The association of business rules draws on the same data as the mined key performance indicators. By mining of the before mentioned dataset and using the key performance indicators as input, variables for mining data result in associated business rules. This research tries to find relations between the available key performance indicators and business rules, resulting in knowledge expansion.

The contribution of this research is the development of a method that 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.

2. RESEARCH DESIGN

This chapter portrays an introduction and an overview of the research design, consisting of a detailed description of the problem statement, research objectives, questions, environment, relevance, strategy and method. Furthermore, this chapter provides a process deliverable diagram for this research, showing the deliverables related to each of the research processes.

2.1. PROBLEM STATEMENT

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.

Gathering knowledge from these pools of data, entails analyzing the historical data available in the databases. This research proposes a method to select relevant cases for connecting information, providing knowledge to customers in the form of key performance indicators and associating business rules. The research proposes a method to extract key performance indicators and business rules from data. Next to the extraction, the method describes constraining indicators by using business rules that provide lower and upper thresholds for these indicators. The extraction and generation of business rules and key performance indicators and presenting these in human readable formats enable organizations to make rational decisions and increase performance.

Summarizing the problem leads to the following problem statement:

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

2.2. RESEARCH OBJECTIVES

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 above mentioned sums up the activities in finding a solution for the problem statement as defined in the previous section.

2.3. Research questions

The previous sections concerning the problem statement and research objectives describe the initiation of this research, as well as the objectives this research tries to accomplish. This section describes the main research question(s) formulated, by keeping in mind the problem statement and research objectives. Firstly, the main research question is introduced. Secondly, the supporting sub research questions are introduced, which have to be answered, in order to provide an exact answer to the main research question.

2.3.1. Main research question

The main research question for this research originates from the problem statement as mentioned in section 2.1.

"How can historical data about employee and organizational performance be transformed into business rules that define the effects of fluctuations in performance indicators within a customer service center?"

2.3.2. Sub research questions

Below are the sub research questions this research tries to answer. For the purpose of identification, the research questions have been categorized to research topics applicable to this research.

2.3.2.1. Business intelligence

The basis of this research lies within the domain of business intelligence, which is defined as the creation of knowledge within an organization, derived from historical data in the form of raw data. Knowledge accumulated from researching this subject will consist of input for extracting, transforming and loading of data. Within the domain of business intelligence, this research will concentrate on business analytics, which in this research is defined as techniques aimed at obtaining new insights and understanding old data. These new insights deliver key aspects for the creation of a method, for the extraction of information in the form of key performance indicators and business rules. The following questions have been defined for the domain of business intelligence:

i. How can (business) analytics assist with the extraction, transformation and loading of data arranged within the area of business intelligence?
ii. How can we define patterns from data using business analytics?

2.3.2.2. Data mining

Researching data mining delivers knowledge consisting of different data analysis methods, which exist within the area of data mining. In order to obtain a clear overview of performance indicators available within databases and performance defining business rules, methods from the field of data mining are selected and applied for data analysis. Next to providing input for the selection of relevant performance indicators, data analysis through data mining provides an initial input for the definition of business rules.

The exploration of data mining leads to the following research questions:

i. Which techniques are applicable for mining historical data?
ii. How can we select performance indicators from historical data using data mining?

2.3.2.3. Key performance indicators

This keyword provides knowledge to select key performance indicators from a great pool of performance indicators, the requirements to turn a performance indicator into a key performance indicator. In addition, to construct Specific, Measurable, Achievable, Relevant and Time phased performance indicators. To incorporate key performance indicators and needed interesting parts, such as performance metrics applicable to this research, the following research questions need an answer:

i. What are the required elements for a (key) performance indicator?
ii. What makes an indicator a key performance indicator?
iii. Which (key) performance indicators can be derived from the historical data available in the data available in this research?

2.3.2.4. Business rules

The definition of SMART key performance indicators is supported by defining business rules, which help to define the performance (measures or thresholds) of a key performance indicator. Business rules constrain processes, as well as guide in the process of decision making; therefore, the following research questions have to be answered:

i. What are the required elements for a business rule?
ii. How should business rules be presented to an analyst or system?
iii. Which business rules can be derived and aligned to the previously mined data?

2.4. Research relevance

Scientific research should serve the public and the research community. Below are the contributions this research tries to realize.

2.4.1. Scientific relevance

The scientific contribution will consist of a generalizable method, from which organizations can derive key performance indicators from raw, available data in databases. The method will provide organizations with aspects needed for extraction of performance indicators. Organizations will be provided a guide to select the performance indicators, which will have a long-term (strategic) influence on performance. Furthermore, the research will provide input for selecting the associated business rules to a key performance indicator based on data mining. Business analysts will review the method, in order to perform evaluation.

2.4.2. Business relevance

When looking at the business relevance of the research, performance indicators are well spread within databases available to organizations. Business relevance lies within enabling organizations to provide customers with answers to questions not based on guesses and averages, but on analyzed rational data. By providing customers with answers to questions they cannot answer themselves, the results will continue to evolve into their systems, enabling other large organizations to profit from the performed research.

The research focuses on providing organizations with a way to increase further performance in these times. Social benefits could entail from the following:

- Transparency concerning complaints/performance of products
- Better pricing models for products
- Faster response from servicing organizations.

3. Methodology

This section describes the various research strategies this research is based on, knowing: qualitative or explorative analysis and design research.

3.1. Research methods

Below is the model view for this research, it is created following the approach as defined by Verschuren & Doorewaard (2000).

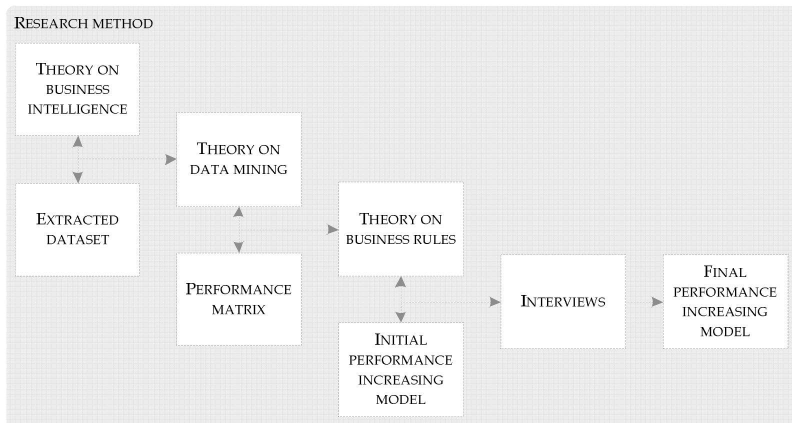

Figure 2: Research model

Each rectangle within the research model identifies a research object(ive) that needs to be obtained to answer the main research question and compliance to the research objectives as drawn in section 2.2. The model is created by taking the right to left approach, starting with the final or end result and analyzing the in-process steps, which have to be taken to reach this result. This will contrast the order of the steps, which will be fulfilled within the research itself and are taken from left to right attributing to the end result.

Firstly, theory on research topics and a case study consisting of combining key performance indicators and business rules result in a performance matrix. Essentially, the matrix aligns key performance indicators and business rules. Secondly, the matrix and an inventory of current (key) performance indicators are analyzed and combined to an initial key performance matrix model, following the method which describes the extraction of key performance indicators and associated business rules. Thirdly, this key performance matrix model is validated by interviews with project managers, resulting in a final result: the final key performance matrix.

3.2. Qualitative analysis

Hart & Boeije (2006) consider qualitative analysis as a means to take part in the situation at hand, by using in-depth/open interviews or case studies. In this case, the researcher is considered a research tool. When measuring the quality of the analysis, reflection on activities, as well as the role the researcher took when conducting the research is important. Rossman & Rallis (2003) describe five characteristics of qualitative research: deeming qualitative research to (a) be naturalistic, (b) drawing on multiple methods that respect the humanity of participants in the study, (c) focusing on context, (d) being emergent and evolving, as well as (e) being fundamentally interpretive.

Corbin & Strauss (1990) note that qualitative methods allow for acquiring a better understanding about any phenomenon, which has not (yet) been researched (often). Qualitative methods also allow gaining new perspectives about subjects, which have a broad research basis or gain more in-depth information concerning the subject of research. Since little research is performed on the extraction of key performance indicators and associated business rules from raw data, plus an excess of information is available on the extraction of information from raw data, both examples speak in favor of qualitative analysis.

Genre	Main strategy	Focus of inquiry
Individual lived experience	In depth interviews	Individuals
Society and culture	Case study	Groups or organizations
Language and communication	Microanalysis or text analysis	Speech events and interactions

TABLE 1: QUALITATIVE GENRE AND OVERALL STRATEGY (ROSSMAN & RALLIS, 2003)

Table 1 lists the strategies applicable within the field of qualitative analysis. From these a case study strategy will be selected. Since the focus of inquiry concerns the whole organization, in-depth interviews will be used as a means of evaluation with regard to the external validity of the research.

3.2.1. LITERATURE STUDY

A literature research is conducted, resulting in hypotheses and possible answers to several proposed research questions, concentrating on findings by key researchers. This is shown in the research model (figure 2) as theory on research topics. Hart & Boeije (2006) propose a systematic approach to literature research, initially describing the subject and performing a query on top-level literature, resulting in information which acts as input for defining an in-depth search-plan (consisting of a detailed description, which information to take from what location and the most important aspects). This in-depth search-plan and results from data analysis will serve as an input for the construction of a method for alligning business rules and key performance indicators. The literature study will provide answers and insights in the research topics described in section 2.3.2 for the means of: business intelligence, data mining, key performance indicators and business rules.

3.2.2. Case study

As detailed in section 3.1, the 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 the influence can be measured and presented in the form of business rules. In order to create a methodological sound case study, this section will discuss the case study (set-up) for this research. 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. This research uses the following definition for a case study: a case study is a research strategy comprising of an all-encompassing method, which covers the logic of design, data collection techniques and specific approaches to data analysis in order to investigate an individual unit.

Several misunderstandings exist among scholars on case study research; the following five are most occurring:

i. General, theoretical (context-independent) knowledge is more valuable than concrete, practical (context-dependent) knowledge.
ii. One cannot generalize on the basis of an individual case; therefore, the case study cannot contribute to scientific development.
iii. The case study is most useful for generating hypotheses; that is, in the first stage of a total research process, whereas other methods are more suitable for hypotheses testing and theory building.
iv. The case study contains a bias toward verification, that is, a tendency to confirm the researcher's preconceived notions.
v. It is often difficult to summarize and develop general propositions and theories on the basis of specific case studies. (Flyvbjerg, 2006)

By following the approach to the design of case studies as defined by Yin (2003), several of the above misunderstandings are dealt with. Yin discusses the use of a case study design to overcome the concerns regarding generalization, which in the case of case study research can be regarded as generalization by theoretical propositions. The major goal for conducting this case study research is expanding and generalizing theories available, also described as analytic generalization. By meticulously defining the research set-up and keeping in mind the construct validity, internal validity, external validity and reliability of the case study, subjective judgments on determining the data and its results can be ruled out.The case study set-up described in this section is the basis for a single case exploratory case study, which is determined by Yin (2003) as generalizable to theoretical propositions and not to populations or universes. In the following sections the elements of the case study set-up are discussed.

3.2.2.1. Research question(s)

Yin (2003) and Perry, Sim, & Easterbrook (2004) suggest a case study should start out with a research question and the case study should systematically collect and analyze data to answer the initial research question. A set of basic principles lie at the basis for the definition of the case study carried out during the writing of this thesis. First an analysis of the form of the research question for this research is needed:

"How can historical data about employee and organizational performance be transformed into business rules that define the effects of fluctuations in performance indicators?"

Yin describes in his work on case study research that the form of the research question often allows to determine the research strategy, from his research and the above research question we can determine that the research question in the form of "how" allows for the case study to be a satisfactory method. The choice for case study research is also satisfied by the fact that the set of elements as proposed in the research question is not under the influence of the investigator. In other words no control is issued about the events taking place within the organization.

3.2.2.2. Possible propositions

Since an exploratory case study is due, propositions are not required or satisfactory for judging the case study's success; this is the case in which a topic is the subject of exploration and also existing in surveys, experiments and other research strategies (Yin, 2003). In order to determine the success of the case study we need to find an answer to the research question posed. By describing the extraction and transformation of data to business rules and key performance indicators, one is able to provide an explanation for the effects measured.

3.2.2.3. Unit of analysis

Determining the unit of analysis is performed by analyzing a subsection of the research question, in this case the customer service center. However, in determining the unit of analysis, it comes forward that the customer service center is suitable for analysis. In fact, the most interesting and logical step would be to look at the organizational cooperation between the customer service center and a large Dutch telecommunications provider. Thus, in this case study the unit of analysis is the organizational cooperation of a customer service center with the telecommunications provider. The cooperation exists in the execution of an operational campaign for the telecommunications provider, this campaign is aimed at consumers which have an existing connection which can be internet or telephone. The campaign is targeted at providing customers with solutions based upon technical or informational inquiries by these customers.

3.2.2.4. Data collection

The data for this case study is two-sided; performance results will be provided by the telecommunications provider. The data provided is measured by systems and not subject to bias. By making a comparison between organizational performance and performance on an employee level, data concerning the performance of the employee is gathered by a team of data collectors situated at the customer service center. The results are a database of performance by the organization's employees and track the performance of an employee on an inquiry of the customer.

3.2.2.5. Logic linking of data to propositions

From the data one should be able to find independent variables influencing dependent variables. By pattern matching independent variables and concluding whether their change effects the dependent variables. The data is linked to the proposition, which in this case is burdened with how the dependent variables are influenced by independent ones and how these effects can be described and constrained in the future.

3.2.2.6. Criteria for interpreting findings

Interpreting the findings from the case study is performed by determining if the results suffice 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.

3.2.2.7. Theory development

Theory development, in this case, is concerned with discussing how performance indicators can be derived from historical data. The case study will show how key performance indicators can be extracted successfully and formed into business rules that constrain the indicator. The case study will show the data collected from various sources is connected by using the Rule Extraction Matrix method. By using the REM-method changes in employee performance, influence organizational performance. These influences will be translated in performance indicators and business rules.

3.2.2.8. Case study protocol

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 which monitor service requests on a project using a five-point Likert scale and rank a monitored service request based upon thirteen different quality related performance measurements. Quality employees are presented with a random selection of service requests that are available for ranking, 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. Results are reviewed monthly and outliers are selected and reviewed with the quality employee responsible for the ranking. Furthermore, a quality employee is summoned for a quarterly training to increase the reliability of the results.

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. One of the biggest challenges lies in the fact that not every service request can be rated completely due to various reasons. One of them being a disconnection by customer or technical issues, or the customer is sending his inquiry to the wrong department. Service requests that are not ranked completely will not be included in the analysis of the data.

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

3.2.2.9. Validity

As stated in section 3.2.2, a common misunderstanding upon case studies is occurring, case studies contain a bias toward verification, in order to overcome this sections addresses the validation of case study.

3.2.2.10. Construct validity

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 by closely defining the steps of the case study with regard to data sources as well as data location(s). As shown in section 3.2.2.4 and

3.2.2.8, the data is collected from multiple sources and the process of data collection and data analysis is described in detail in chapter 6.

3.2.2.11. Internal validity

Section 3.2.2 detailed the case study being an exploratory case study due to the research question being of the "how"-form. 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.

3.2.2.12. External validity

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 Rule Extraction Matrix 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 (Rule Extraction Matrix-method) is used as the template, by which the empirical results of the case study are compared.

3.2.2.13. Reliability

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. Both techniques will be included within this research. A brief description of the case study protocol can be found in section 3.2.2.8. The case study database and other materials associated with the case study have been provided within the appendixes of the research document, these include the data structures and metadata available for the case study, interview questions, the interview responses and a mock-up dashboard to present the results to the end-user.

This section will chronologically describe the steps within the case study below. Progress of the case study will be documented “as-is” following the steps of the proposed method in detail within chapter 6.

- i. Construct case study design
  - a. Research question
  - b. Determine unit of analysis
  - c. Define propositions
  - d. Theory development

- ii. Data collection
  - a. Extract data from systems
  - b. Data collection by employees
    - i. Verification of the data (outlier detection)
    - ii. Quarterly training

- iii. Data analysis
  - a. Linking data to propositions
    - i. Determine correlations between data
    - ii. Set-up performance indicators and business rules

  - b. Verify findings and theory
    - i. Verify findings matching criteria
    - ii. Verify findings by interviews
      1. Prepare data for interpretation
      2. Set-up interview questions
      3. Interview respondents
      4. Summarize results

- iv. Finalize case study report
  - a. Report findings
  - b. Present

3.2.3. Interviews

Semi structured in-depth interviews are conducted as means of data validation finalizing the performance matrix and transforming the result into a key performance matrix model; therefore, enabling organizations to use this model to support decisions by tangible knowledge.

3.3. Design research

Since little research has been performed on combining key performance indicators with business rules, the development of a model as shown in figure 2 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.

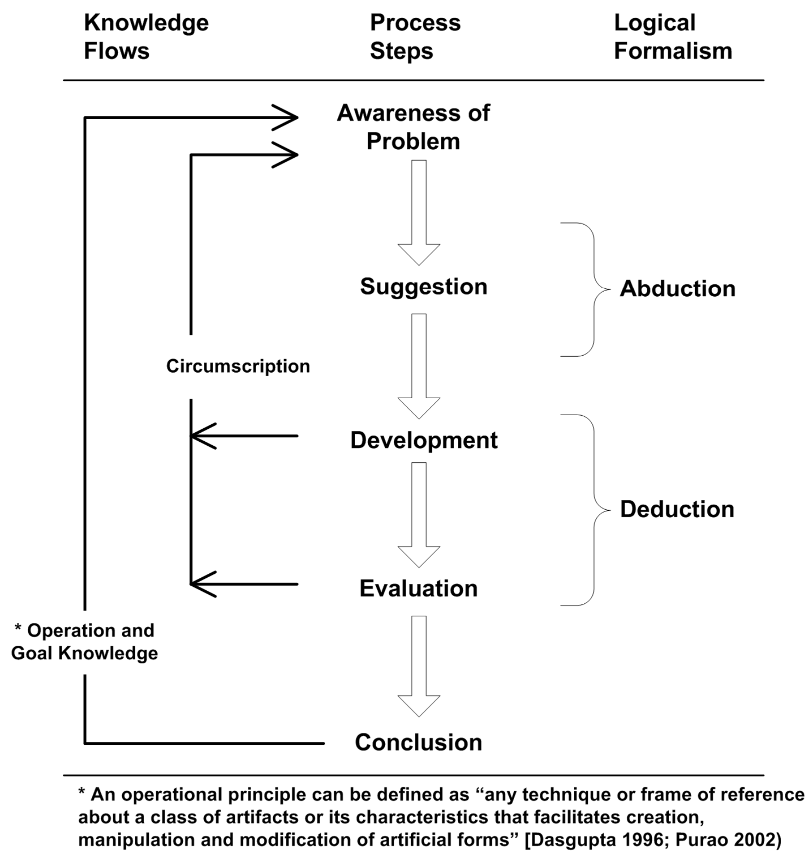

FIGURE 3: REASONING IN THE DESIGN CYCLE (TAKEDA, TOMIYAMA, YOSHIKAWA, & VEERKAMP, 1990)

The steps in developing a method will take the approach of design science as shown in Reasoning in the design cycle figure 3 taken from Takeda, Tomiyama, Yoshikawa, & Veerkamp (1990). The first process step in the design cycle is the awareness of the problem, as described in the section on the problem statement, this research addresses the following 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. This research constructs knowledge from historical data by aligning or associating business rules to key performance indicators, which can both be found in historical data available in performance databases. The third step portrays the creation of an artifact, based on the knowledge gained from the completion of the second step, which is deemed to be the key Rule Extraction Matrix Method, as shown in figure 2 and its accompanying description. The fourth step within the design cycle shows the evaluation phase, in which the key performance matrix (or artefact) is evaluated and eventually elaborated on by the fifth step, following a normal scientific research approach.

3.3.1. Development of a model

As described in figure 2 and the section on design research, this research incorporates development of a model. The Rule Extraction Matrix model combines existing knowledge plus knowledge gained from field work and case study to provide organizations with a solution for gaining knowledge (in the form of key performance indicators and associated business rules).

3.4. Selected sources

Table 2 below shows the sources of information that this research will incorporate, in answering the research questions, below the research questions have been labeled by subject and corresponding numbering as seen in section 2.3.2.

Source	Disclosure	Research questions/validation
Books, literature, papers, white-papers	Literature study	Business Intelligence: i,ii Data Mining: i,ii Key Performance Indicators: ii Business Rules: i,ii
Organizational data	Case study	Key Performance Indicators: i,ii,iii Business Rules: i,ii,iii
Project managers, business analysts	Interviews	Key Performance Indicators: ii Business Rules: ii

Table 2: Selected sources

3.4.1.1. Literature study

The literature study will focus on books, literature found by scanning websites, including papers, relevant articles and white papers eventually resulting in a long list. By selecting relevant literature following the qualitative systematic approach, as described by Hart & Boeije (2006), the literature research results in a firm research basis to:

i. Propose a model
ii. Select cases for the case study and execute the case study
iii. Prepare for interviews.

The long list of literature results in an in-depth search-plan on the applicable research topics, mentioned in section 2.3.2 and 3.4. The in-depth search-plan can be found in Appendix A: in-depth search-plan. Furthermore, the literature study will also be used to provide the basis for conclusions and further discussion.

3.4.1.2. Case study

As shown in the previous section, the case study is based on results obtained by selecting relevant information in the literature study. The execution of the case study is based on knowledge, gained from the literature, research on data and on case selection. In addition to knowledge, a selection is determined by relevancy of databases and applicable access to these sources, provided by the approval of service providers.

3.4.1.3. Interviews

Following the approach of Hart & Boeije (2006), there are four elements of structuring an interview and the amount of structure applied on four levels. These levels determine whether an interview is an unstructured or free interview, semi- or half-structured interview or a standardized or structured interview. The four levels are defined as:

i. Contents of questions is more or less fixed
ii. The presentation of questions is more or less fixed
iii. The order of questions is (not) variable
iv. Possible answers have (not) been provided.

The way of interviewing applicable to this research is the semi-structured interview, by a solid definition and presentation of questions, this interview abides to the open or qualitative interview as defined by Merriam(1998).

Below is the planning for an interview:

i. Contact a respondent for a face to face interview
ii. Explain why the respondent is interviewed and how the data will be used
iii. Set-up a meeting between the interviewer and interviewee (respondent)
iv. Provide an introduction concerning the interview
v. Introduce the role of the interviewer
vi. Introduce the subject of the interview
vii. Execute the interview (audio recording as a means of backup)
viii. Ask for confirmation (brief summary of noted answers) to the answers
ix. Thank the respondent for his/her time and attention
x. Create a summary of the interview
xi. Present the summary to the respondent
xii. Verify the interview findings with the respondent
xiii. Document the findings within the case study database.

3.5. Process deliverable diagram

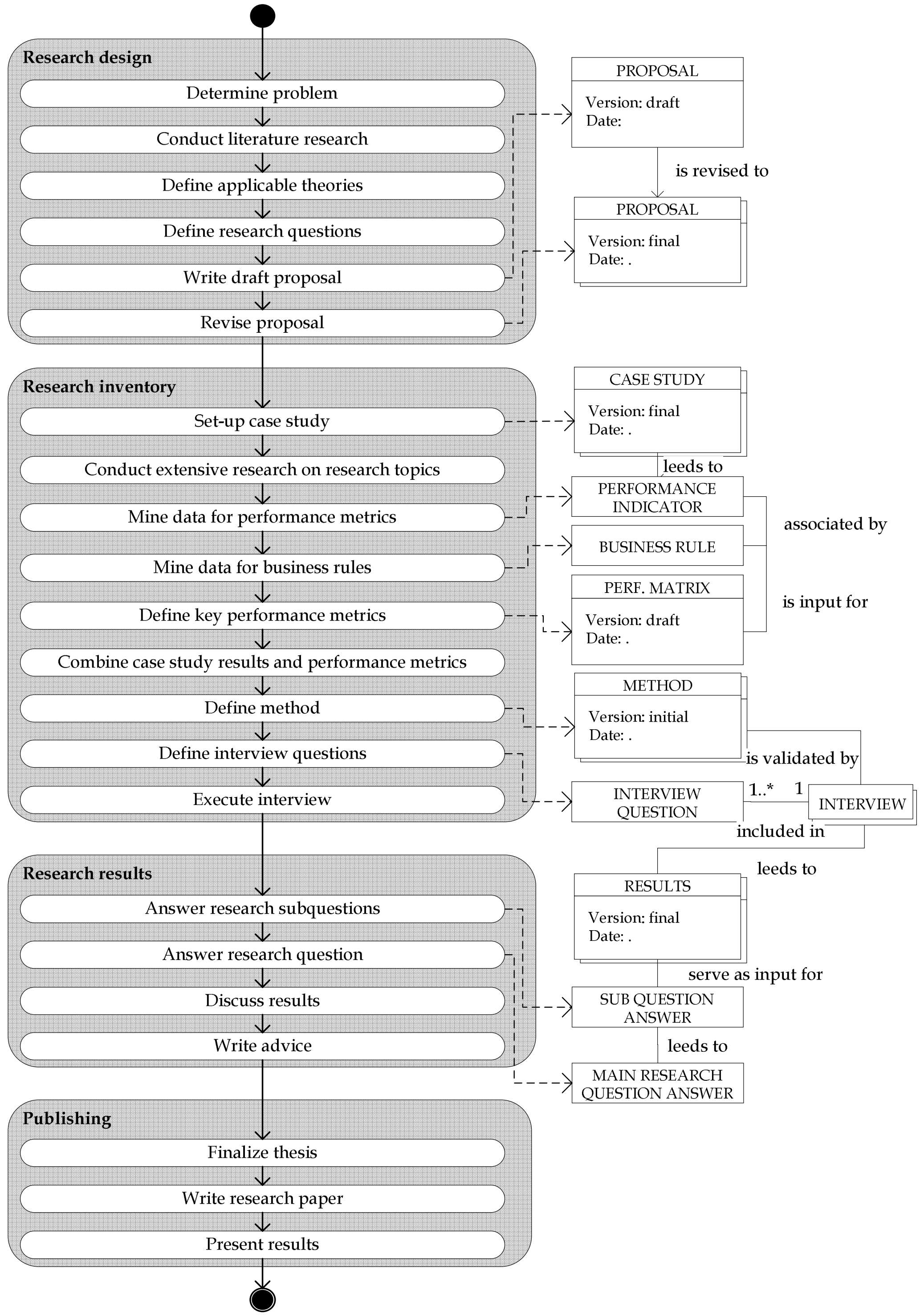

Figure 4: Process deliverable diagram

4. THEORETICAL BACKGROUND

The sections below describe the domains/topics, which are applicable to this research and make explicit what kind of knowledge the research uses to form a method for the combination of key performance indicators and business rules. Moreover, the theoretical background offers suggestions for selecting relevant cases from a pool of service providers, which are incorporated in the case study. Furthermore, the theoretical background constructs a foundation for the interviews which are mentioned in section 3.2 on qualitative analysis. Next to domain selection and knowledge increase, the theoretical background also highlights the research relevance of this research and provides a basis for answering research questions as stated in section 2.3.

4.1. BUSINESS INTELLIGENCE

The focus of this research lies within key performance indicators and business rules; an overall umbrella definition for the two main topics within this research is business intelligence. Business intelligence is a well-known topic for several 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, Kreulen, Krishna, & Spangler, 2002).

Research on business intelligence originates from the term as introduced in 1989, by Howard Dressner, regarding business intelligence to describe 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).

As 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 5 and figure 6 listed below.

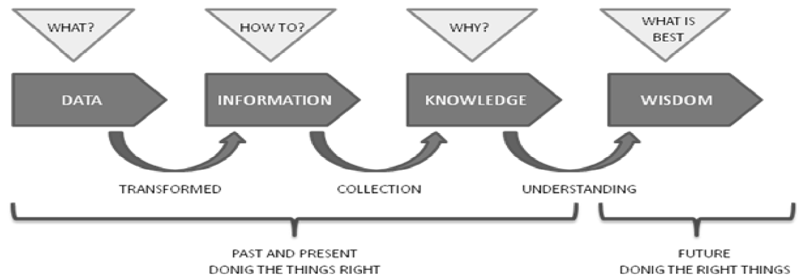

FIGURE 5: DATA TO WISDOM TRAVERSAL (FUGANTE, 2008)

Data to wisdom traversal known from the knowledge hierarchy, also displayed as the infamous knowledge pyramid, is shown in figure 5, is found in literature on business intelligence (Rowley, 2007) (Piatetsky-Shapiro, 1991) as shown in figure 6 below. Based on both figures and the definition devised by Howard Dressner, the (bounded) definition used in this research for business intelligence is the organizations' ability to construct knowledge in the form of decisions from structured and unstructured data available in (legacy) systems.

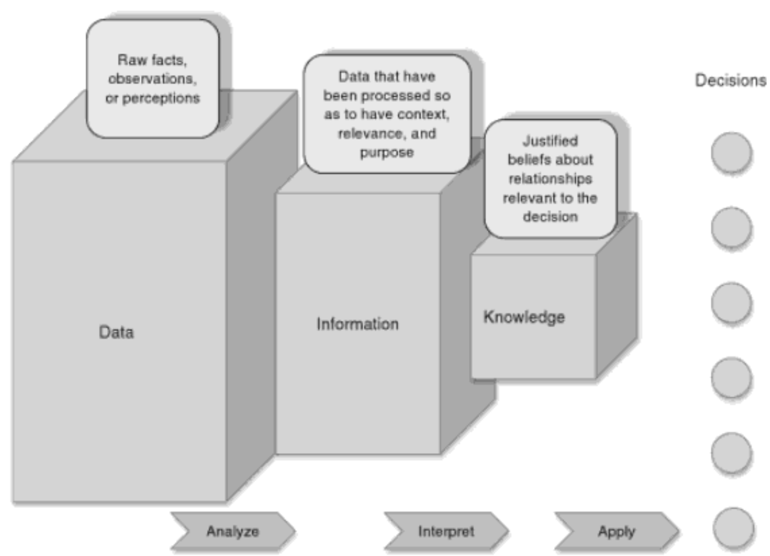

FIGURE 6: DATA TO WISDOM TRAVERSAL IN THE AREA OF BUSINESS INTELLIGENCE (ROWLEY, 2007)

As shown in the definitions and figures listed above business intelligence relies on business data in the form of raw facts, observations and or perceptions that can be analyzed for making decisions. To form decisions data has to be transformed or processed into information, in essential data combined with meta-data (also known as context); this step is often handled by data warehouses. The last step in the creation of decisions from raw data is transforming the information into knowledge. Knowledge-creation is the process of determining relations between information, such as a key performance indicator known as customer satisfaction influencing a financial indicator (e.g. the net sales rate of an organization).

4.1.1. Managing data

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. Business intelligence uses sophisticated techniques available in modern databases management systems, to build large data warehouses and extract business advantage from these large pools of data using data mining techniques. 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 listed in figure 6. 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):

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

The above is in close relation with knowledge management, with regard to its strategy of identifying, creating, distributing and enabling adoption of insights and experiences; insights comprise knowledge embodied in individuals or in the case of this research embedded in organizational processes and practice. Knowledge management affects the conversion from tacit to explicit knowledge and vice versa (Lytras, Russ, Maier, & Naeve, 2008). Essentially, it is applicable to the research problem, where tacit knowledge is used to provide answers to problems, but the process of answering is not always sufficient or repeatable.

4.1.2. Developments

Two recent developments in the area of business intelligence are closely related to this research; these are business performance measurement and business activity monitoring. A third term attributed to Negash and Gray(2008) is business intelligence for the masses, by stating that business intelligence evolves from being a typical analyst business to the development that people throughout the organization use business intelligence to improve the execution of their tasks and processes. Business performance measurement and business activity monitoring were brought to the attention of researchers in the 2000s, the main fundamental ground of both theories, is finding how an organization performs by combining software, processes and measurements of business success. Where business performance measurement fundaments lie in the long run, business activity monitoring takes the real time component into account.

Business performance measurement - If and when an organization looks at its business performance, an organization will undoubtedly come across business performance measurement. This enables organizations to combine recent performance measures, such as the balanced scorecard and key performance indicators. Focus within the area of business performance measurement lies within the consolidation of multiple measurement instruments. As described in the previous paragraph, this is not performed in real-time situational analysis,

Business activity monitoring – Focus within business activity monitoring lies with high volume production, where organizations need to manage every aspect of production, in terms of business activity outputs and key performance indicators in real-time or close to real-time. Combining measurements and corrective actions that are measurable from a business activity monitoring perspective is referred to as the observation-orientation-decision-action loop (Negash & Grey, 2008), or the model of quality management (Deming, 1986): plan, do, check and act.

Recent developments within the area of business intelligence contribute to this research regarding score-carding mechanisms, as seen within the domain of business performance measurement. Score-carding mechanisms serve as an input for the definition of key performance indicators and measure the effectiveness of decisions. This is also seen within the domain of business activity monitoring which contributes to the effectiveness of the extracted information in the form of decisions.

4.1.3. Users and appliances

Google searches for "business intelligence for the masses" or "business intelligence for all" lead to a widespread amount of hits, ranging from blogs to organizations. As stated in the section on recent developments within the area of business intelligence, a shift from analytic users to business intelligence for the masses is occurring. This is noticeable from the huge amounts of information available in various sources. From these sources, Internet serves as the largest information source. Business intelligence is a domain attributing to many appliances such as:

i. Querying and reporting
ii. Real-time analysis
iii. Forecasting
iv. Decision support systems
v. Dashboards and scorecards input.

The above is also noticeable on a user level, with users of business intelligence techniques ranging from the employee, who needs to know its current sales target coherence, to a chief executive officer demanding forecasts on the financial climate of an organization.

From a business intelligence perspective, users can be divided into casual users and power users. These business users obtain insights from business intelligence tools and techniques. Supported by business intelligence, these users will provide an organization with the right decisions and actions to gain tactical and strategic advantages and eventually reach their target goals. Organizations, on a whole, often come to the conclusion that analyzing data is not easy without proper resources such as tooling, prepared data and the access to business analysts.

For organizations to successfully adopt and maximize the potential of business intelligence, end users should be provided with a solution, which is tailored towards their needs and characteristics. By identification and classification of (end-)users, solutions in terms of business intelligence tooling and functionality matching the expectations of end-users is achieved, resulting in higher (end-)user satisfaction rates and easier adoption of business intelligence solutions(Tijssen, Spruit, van de Ridder, & Raaij, 2010).

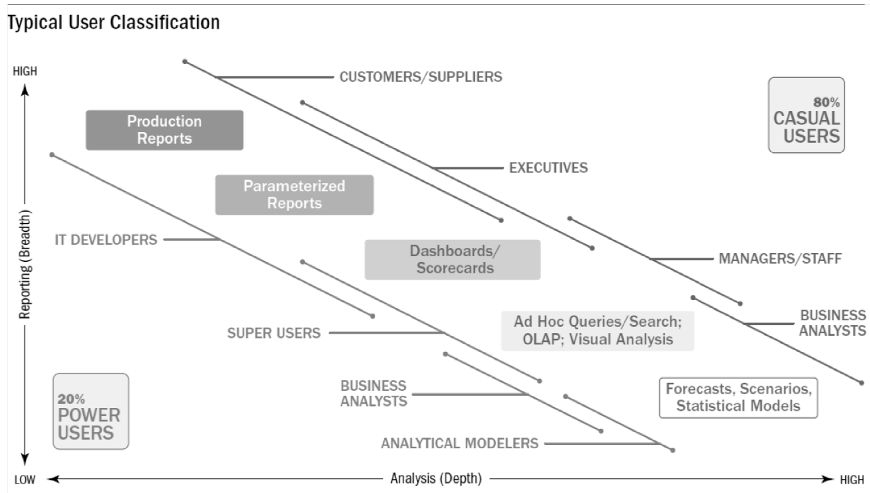

FIGURE 7: CLASSIFICATION OF BUSINESS INTELLIGENCE USERS (ECKERSON, 2009)

Figure 7 displays a user classification into power users and casual users; moreover, the graph shows the partitioning between the classifications of users, stating that casual users, also referred to as consumers (of data), outnumber the power users, referred to as producers (of data), by four to one. However, the outnumbering is relative since by producing the business intelligence content, designing roadmaps and selection of tools, these power users have an oversized impact on business intelligence environments within organizations (Eckerson, 2009).

This research will mainly focus on the area of business analysts, with regard to power users and managers/staff on a casual user's perspective, since these areas both cover and overlap the area of key performance indicators, decisions and global scope of this research. Eckerson (2009) describes the MAD framework for business intelligence environments, categorizing functionality to end-users.

Functionality	Users	Appliances
Monitor	Executives/Managers	Graphical key performance indicators
Analyze	Managers/Analysts	Dimensional views and filters
Drill	Analysts/Workers	Operational queries and reports

TABLE 3: MAD FRAMEWORK

The MAD framework is shown in table 3 above, it displays the roles and appliances associated with business intelligence functionality. With regard to the framework, analysts are known to perform three major tasks; gathering data, analyzing data and presenting data to management or executives. Current technology, within the area of business intelligence, enables business analysts to optimize the processes of gathering, analyzing and presenting information without interfering with the global enterprise standards.

4.1.4. Shortcomings

In short, business intelligence literature displays no effective mechanism to translate business information in the form of decision support to end users/analysts or systems. One effective way to display information in the form of decisions is applying business logic, in the form of key performance indicators and business rules, to help conceptualize the impact of key performance indicators and offer a clear overview of a key performance indicator to users of various business intelligence tools or toolkits.

Return on investment – Studies conducted in 2003 show returns on business intelligence ranging from 17% up to 2000% averaging 457% (Morris, 2003) (Darrow, 2003) whilst not showing a correlation between business intelligence budget and return on investment. Business intelligence implementation and upkeep costs are high; this in combination with difficulty computing anticipated returns makes organizations reluctant towards buying business intelligence solutions (Negash & Grey, 2008).

4.2. Data mining

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 the domain of data mining and business intelligence are 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, Piatetsky-Shapiro, & Smyth, 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 modelling 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, Piatetsky-Shapiro, & Smyth, 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, Piatetsky-Shapiro, & Smyth, 1996) (Cooley, Mobasher, & Srivastava, 1997).

Approaching data on a huge scale is a necessity to enable organizations to gain competitive advantages, increase efficiency and offer better customer experiences all relying on input from processed data (information). Data mining enables the analysis of a whole set of data or in case of huge datasets and little resources, subsets of this data to make predictions for the whole set and future, however; lack of available resources is becoming a smaller risk with the growth in computational power, resources available within organizations and the efficiency of data mining algorithms.

Knowledge discovery from databases – This act is described by Fayyad, Piatetsky-Shapiro, & Smyth (1996) and shows similarities to the process of extracting, transforming and loading of data as displayed within the section on business intelligence. Knowledge discovery from databases covers nine steps knowing:

i. Developing an understanding of the application domain/relevant prior knowledge and identifying the goal from a customer's point of view
ii. Creating a target set of data through selection or focusing on a subset of data samples/variables
iii. Data cleaning/pre-processing through the removal of noise
iv. Data reduction and projection
v. Matching the goals determined in step i to a particular data mining method/algorithm
vi. Exploratory analysis/model and hypothesis selection
vii. Actual data mining
viii. Interpreting the results that follow from the previous step
ix. Taking appropriative action on the results discovered from the previous process step(s).

Step ix. involves the data to wisdom transformation, through gaining information from raw data, knowledge discovery from databases contributes to the information available within an organization and ultimately provides input for the data to wisdom traversal as described in section 4.1.

CRISPDM – An addition to knowledge discovery from databases is the cross industry standard process for data mining (CRISPDM), shown in figure 8 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 new 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).

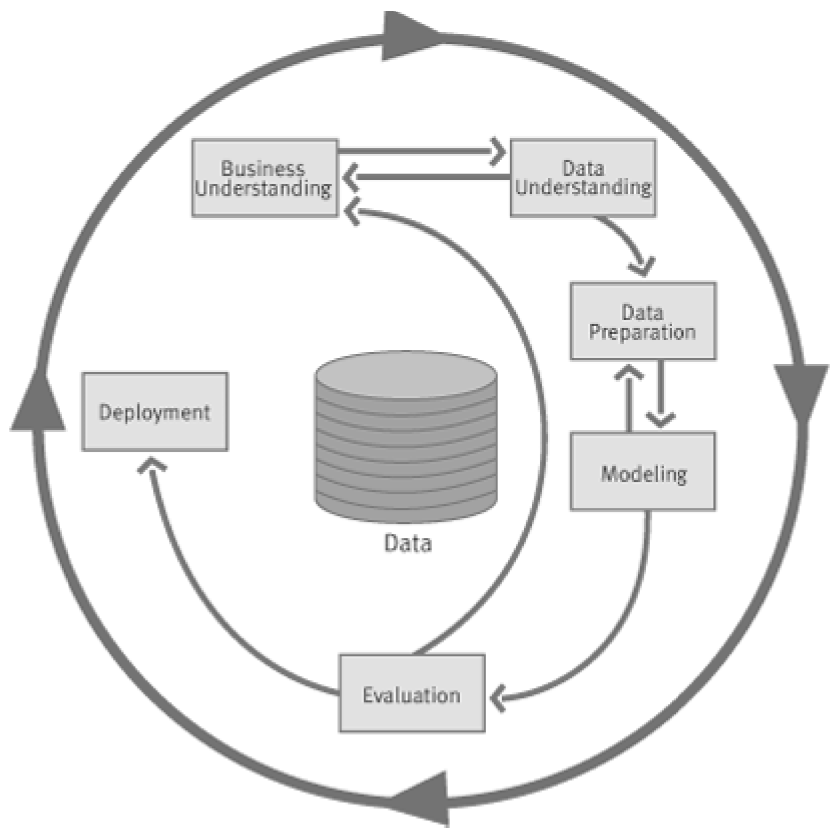

FIGURE 8: CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING MODEL (CHAPMAN, ET AL., 2000)

BUSINESS UNDERSTANDING – Charting the project objectives and project requirements and forming this into problem definitions suitable for data mining and a plan to achieve objectives.

Data understanding – Data collection and judge the quality of the collected data to determine possible problems and first insights.

Data preparation – Construction of the final dataset that will be fed to the modelling tool(s) from the raw data available.

Modelling – Selection of techniques and calibration of parameterized techniques.

Evaluation – Review steps that helped form the model and determine the relation of the model towards the business objectives.

Deployment – Presentation of the data to the customer.

This research will focus on combining the model provided by Fayyad, Piatetsky-Shapiro, & Smyth (1996) with the additional processing steps detailed in the model by Chapman et al. (2000). The process of combining the two methods is comparable to the research approach chosen by Vleugel, Spruit, & van Daal (2010).

4.2.1. Data mining methods

The process of data mining is the application of specific algorithms/techniques for extracting patterns from data. Methods within the data mining domain are concerned with the primary goals of prediction and descriptions. Prediction models entail the usage of available data to predict the future in terms of possible outcomes. Description models enable the display of patterns, which explaini the past using the available data and translating the data in to human readable formats through statistics and visualization techniques. Prediction and description are not exclusive, predictive models are capable of being descriptive and vice versa.

Various well known models applicable to the descriptive and predictive properties of data mining including examples are displayed below:

4.2.1.1. Description

Description models or exploration models allow the past to be displayed by graphical or statistical techniques. This exploration allows analysts to determine where the focus for further analysis should be. Description can be categorized in the models below:

Univariate – Univariate models analyze variables one by one within a dataset, variables fed to univariate models can be numerical or categorical. Univariate models allow for variables to be transformed from numerical to categorical (binning) or categorical to numerical (encoding). Through univariate models variable transformations can be executed, for example execute analysis on the percentage of employees that achieved a certain target or are in possession of a certain set of skills.

BIVARIATE – Analyzing variables simultaneously, bivariate models explore the concepts of relationships between two variables. The exploration of these variables lies in finding associations and the strengths of these associations or determining differences and the significance of these differences. Through bivariate analysis, it can be determined whether a variable influences another variable, for example: does the amount of salary influence the performance of an employee and if so what is the strength of this relationship.

4.2.1.2. PREDICTION

Prediction models allow for the creation of models that predict the outcome of variables. These predictions can be used to support the decision making processes within an organization. Prediction can be categorized in the models below:

CLASSIFICATION – Classification models involve the learning of a function that predicts a categorical variable based on one or more categorical/numerical variables. By classification models an answer to the following question can be obtained: "Will the employee achieve the targets set by the organization?" By using (categorized) variables (predictors), such as experience and level of training, classification models can answer this question in terms of yes and no.

REGRESSION – Regression models involve the learning of a function, which predicts a numerical variable based on one or more categorical/numerical variables (predictors). An example of the use of classification models is sketched by finding an answer to the following question: "What is the amount of salary needed to attract an employee that has three years of experience and the required certificates/education?" By using data concerning previously hired personnel, the amount of salary that needs to be offered in order can be determined.

CLUSTERING – Clustering models allow a dataset to be divided into groups that allow the members of each group to be as similar (close) to one another and groups to be as dissimilar (far) as possible. The use of clustering allows discovery of undetected relationships in a dataset. Clustering models can be used to answer the following question: "Which age groups excel at technical product trainings?" By using clustering techniques, patterns can emerge from the data enabling an organization to answer the above question.

ASSOCIATION RULES – Association rules models allow analysis of a dataset by finding patterns that describe the relationships between attributes. An example of the application of association rules is finding an answer to the following question: "Given skill A and skill B, what other skills is an employee likely to have?" By using association rules models the answer to this question can be provided by using the data at hand. (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) (Sayad, 2012).

The methods above offer perspective on data mining's many applications for organizations to construct knowledge from historical data. This research is determining prescriptive and descriptive properties of data since a successful prediction of the future needs to be based on a solid base constructed from historical data. This research addresses pattern-retrieval, as well as decision trees combined with association rules at its most fundamental levels, deeming key performance indicators intertwined with business rules.

4.2.2. Selection of data mining algorithms

Selection of data mining algorithms as described by Fayyad et al. (1996) in their research on knowledge discovery in databases needs to be adjusted towards the goal of knowledge discovery, which embodied within this research is the discovery of business rules from historical data. Next to goal influence, the selection of data mining algorithms is highly dependent on the data that is available, for example, financial data stored in a database opposed to error logging from a system; both need different algorithms to deliver the knowledge needed.

Section 4.2.1 outlined the various general data mining methods available within the area of data mining. Selection of data mining algorithms applicable to this research is performed by evaluation of the following three primary components:

i. Model representation
ii. Model evaluation criteria
iii. Search method.

Model representation is considered the language used to offer a detailed description, bound with a certain level of detail on which an analyst is able to comprehend the representational assumptions provided by the model. These representational assumptions might be rooted in a particular method as described above.

Model evaluation criteria cover the quantitative statements regarding the discovered pattern; how well will it fit to the needs, judged by prediction accuracy, novelty, utility and degree of understanding.

Search methods components cover parameter search and model search, parameter search is faced with searching for optimized parameters given the observed data and fixed model representation, whereas model search is faced with offering a loop over the parameter-search method (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

In order to extract business rules and decision trees from the dataset, a selection of the data mining methods (as described in section 4.2.1 needs to be constructed. A successful extraction of business rules requires information about the attributes within the dataset. In terms of data mining methods, analysis of the dataset using univariate models is performed to determine, if the attributes are valid candidates for further analysis. Methods such as binning and encoding need to be applied to the dataset, this allows the data to be fed to the requirements of a specific prediction model. This research concentrates on the extraction of decision trees and translating these to business rules. Decision trees can build classification or regression models using a tree structure, which in turn provide input for association rule models to extract rules from the resulting decision tree. The selection of an algorithm to create the decision tree is highly dependent on the target variable (the one determined or predicted by the model). If a target variable is categorical (for example high or low) a classification algorithm should be chosen, if the target variable is numerical (for example 10,3) a regression algorithm is mandatory. As described above the creation of decision trees and the extraction of rules is highly influenced by the dataset, its possible attributes and their form.

4.2.3. Limitations

While being powerful in essence, data mining products are not self-sufficient; the application of data mining algorithms and structured analysis requires technical and analytical specialists which interpret data and output (Seifert, 2004). Though this is a limitation regarding the application of data mining techniques in general, this research focuses on the highly analytical end-user.

Data mining identifies connections between behaviors and variables within data-sources. However, these connections are not necessarily identified as a causal relationship (Seifert, 2004)(Mazlack, 2003).

4.3. Key performance indicators

Resulting from performance measurement, key performance indicators are a well-known phenomenon in the strategic, tactical and operational levels of an organization. Rockart (1978) first introduced the term key indicators in his work on a new approach to define the chief executive's information needs, since then (key) performance indicators found their way into organizations dashboards, sales meetings enabling organizations to measure results.

The introduction described the key performance indicator as a metric to measure performance in relation to previously obtained results; to do so key performance indicators are often measured at the strategic level. A wide array of key performance indicators can be found in the paper devised by Kaplan & Norton (1992) on the balanced scorecard mechanism, they describe a fast comprehensive view on an organization by measuring indicators at customer, internal innovation and learning and financial perspectives such as cycle time, defect rate, missed deliveries. Kaplan & Norton (1992) state: "what you measure is what you get" and found that no single measure can provide a clear performance target or focus attention on the critical areas of the business, with this in mind, the measurement of key performance indicators and the influence they exert on related key performance indicators enable organizations to keep focus on critical areas of the business.

Through historical performance measurement by monitoring key performance indicators that influence the organizations' results and undertake action where possible to influence the outcome, organizations enable themselves to be in control and optimize financial and operational results. The act of being in control of financial and operational results is described by Cameron & Whetten (1983) as the importance of performance measurement and performance improvement being at the heart of strategic management. While measurement of performance is one of the most important steps in strategic management processes, key performance indicators in their turn are a basic instrument to measure and document performance. The term can be broken down in its three core elements, knowing a key performance indicator is key, when it is fundamental towards the process of gaining competitive advantage and makes or breaks an organizations success or failure. A key performance indicator only relates to performance when it is clearly measured, quantified and easily influenced by actions carried out by the organization. The last element relates to a key performance indicator providing leading information on historical performance and future improvements and/or performance (Nelson, 2010). From this a formal definition for key performance indicators can be derived. This research uses the following formal definition for key performance indicators: The historical and future tracking of measurements, which enables an organization to perform or excel at their most fundamental levels.

4.3.1. Thresholds and monitoring

Enabling organizations to undertake actions based on key performance indicators, is essentially providing them with a clear overview of which performance items need attention. This can be achieved by the expansion of a key performance indicator with thresholds. By adding targets in the form of thresholds to a key performance indicator it becomes manageable. By measurement and keeping a watchful eye on the negative or positive development of a performance metric by thresholds, a key performance indicator becomes a valuable instrument for on-the-fly measurings and undertaking actions.

By visualization of thresholds by incorporating those in dashboards, immediate decisions can be taken to further positively develop these key performance indicators and related key performance indicators, which rely on the outcome of decisions and their effect on other key performance indicators. An example of this can be the average talk time exceeding a certain threshold (480 seconds), this influences not only the key performance indicator average talk time but also negatively influences the amount of calls that can be accepted which in turn has a negative effect on the service level on which other key performance indicators rely, thus closing the loop on a vicious circle.

4.4. Business rules

Morgan (2002) describes business rules as compact statements about the organizations' processes in terms that directly relate to the business using simple, unambiguous language that is accessible to various parties such as a business owner, analyst, and architect. GUIDE (an IBM-oriented industry group) devised the following definition for a business rule: "a statement that defines or constrains some aspect of the business. It is intended to assert business structure or to control or influence the behavior of the business" (Morgan, 2002)(the Business Rules Group, 2000). Describing an entire organization using business rules requires a vast amount of business rules, limits are often imposed on subsystems of an organization and even though these rules are not limited to hundreds, a single rule is small and concentrates on a small part of the business. While concentrating on small parts of organizations a business rule simplifies tasks such as discovery, definition, review and maintenance.

A key performance indicator is in essence a building block such as a stone whereas business rules are the cement which holds a building together and techniques such as data mining and business intelligence could be compared to the plans and analysis constructed by architects. Within this building, events take place and roles are granted to actors with responsibilities. Looking at this from a business architecture point of view the following figure explains the placement of business rules within the organization.

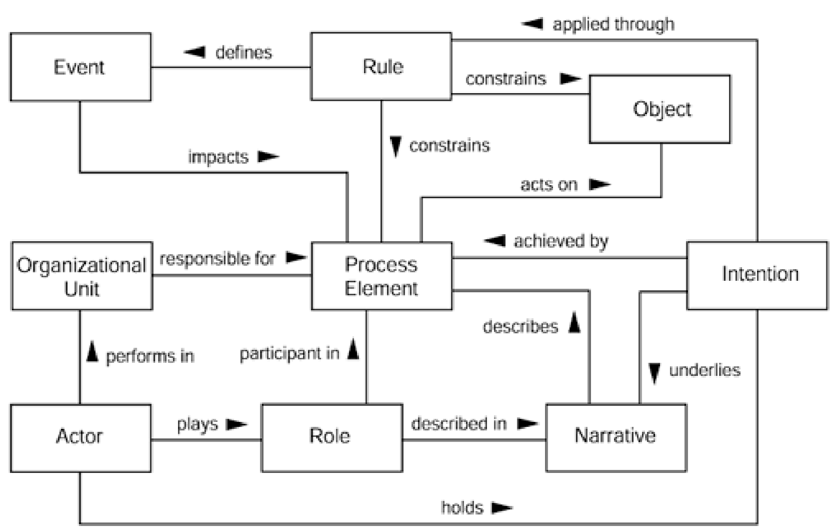

Figure 9: Business architecture (Morgan, 2002)

Figure 9 displays the placement of a rule, stating that it constrains process elements. The above is applicable by regarding a key performance indicator as a process element. Further explaining this situation, while using the already proposed key performance indicators, would lead to the following: an increase in handling times (i) impacts the service level (ii), (i) is the event taking place which can be incorporated within a rule, (ii) is the key performance indicator influenced by enactment of the event.

4.4.1. Presentation

Three levels of presentation exist for displaying rules to various levels, such as analysts, architects and systems.

i. Informal: A service level for a project should be maintained at 80%
ii. Technical: Project.servicelevel >= 80%
iii. Formal: $\frac{Acceptedcallswithintimelimit + Abandonedcallswithintimelimit}{Totalacceptedcalls + Totalabandonedcalls} > 0.8$

The informal level provides natural language statements within a limited range of patterns as displayed through the example above, the technical level combines structured data references, operators and constrained natural language and finally the formal level uses a closely defined syntax, as well as mathematical properties.

Transactions from informal levels to formal levels are often a human activity, which in turn can introduce errors. Informal levels in turn offer organizations a view at decisions and their impacts on the business.

4.4.2. Discovery methods

The discovery of business rules is dependent on data available from a multiple of sources, such as documentation, know-how, automation and business records. Finding rules is bound by several constricts. Firstly, an individual analyst is unlikely to have the breadth of knowledge and experience to understand the business at the right level. Secondly, systematic retrieval is essential to prevent gaps in the information system from appearing. Thirdly, rules need to contribute to the right level of detail preventing too much detail, resulting in difficult implementation and preventing too little detail as an opening for misinterpretation. Finally, large infrastructures need to be taken on as a team, to be effective and spread high workloads.

Table 4 shows the variety of sources opposed to the likeliness of rules retrieval by three discovery methods, knowing: static analysis, interactive analysis, automated analysis. The sources available for this research, applicable techniques and their likeliness for discovering rules are colored grey in the table displayed below.

Source	Static Analysis	Interactive Analysis	Automated Analysis
Documentation	High	Moderate	Unreliable
Tacit know-how	Not applicable	High	Not applicable

Automation systems	Low	Moderate	High
Business records	Depends on source	Low	Depends on source

TABLE 4: APPLICABILITY OF RULE DISCOVERY METHODS (MORGAN, 2002)

4.4.2.1. STATIC ANALYSIS

Static analysis mainly necessitates the careful checking of source documents through electronic copies where possible; this involves the formal categorization by consistent analysis of documents into internal and external documents, internal meaning private to the organization and external being available to other parties and obtained through effort or costs.

Internal sources include but are not limited to:

- Business and marketing plans
- Internally generated reports
- Deliverables and specifications
- Quality manuals and internal directives
- Correspondence and meeting minutes
- Commissioned reports.

Possible external sources are:

- Legislation and standards such as ISO
- Voluntary codes of practice and standard references
- Analytical reports and journals containing publications on relevant cases
- Information provided by partners such as suppliers, customers and/or business partners.

4.4.2.2. INTERACTIVE ANALYSIS

Interactive analysis is the act of instigating discussions by organizing sessions that contain business analysts and business specialists to explore areas, which lack business knowledge in documented form. Interactive analysis can entail the act of interviews or organizing workshops.

4.4.2.3. AUTOMATED RULE DISCOVERY

By the usage of inductive logic and neural networks previously unsuspected patterns from data can be extracted. The extraction of patterns from data has been described in section 4.2 on data mining. One of the main problems in automated rule discovery is the fact business knowledge available within systems is expressed in natural language lacking consistent structure and thus difficult to analyze for data mining systems. The discovered rules should be governed by an employee, preferably an analyst, to ensure the discovery of obsolete and appallingly obvious patterns and rules. (Morgan, 2002)

4.4.3. Rule characteristics

The research combining key performance indicators and business rules complies with the list drawn by Morgan (2002) which states universal characteristics for business rules:

Atomic – Cannot be broken down any further. Business rules should be defined in terms of subunits, but these subunits are not composed of rules, attempting decomposition of these subunits would result in information losses. Rule sets are an exception to the rule meaning that a set of rules might be treated as a super rule, rule sets should be mentioned towards a user of the set to show the composed nature.

Business statements – Unambiguous rules are statements about an aspect of the business, these should display a clear meaning towards the terms they use, facts defined, logic dictated and the use of small sizes for rules reduces the scope for possible misinterpretations.

Declarative – Rules should express a description of goals rather than actions, not how a goal is reached rather, what the goal or business aim entails.

Expressed in natural language – Consistent rules describe the business by the use of natural language; this prevents the need for training, special tools and deems rules extremely flexible. These rules offer a unified description of the case which can be changed in rules that support decisions, rules that can be implemented in systems and mathematical expressions.

Traceable – Coherent rules should be traceable from the back (source of the rule) towards the front (what realizations). Rules should use the same terms to keep in line with the rest of the model when looking at language and definitions.

Structured – By defining patterns for rules, free formats are prevented so business rules can suffice to the previously mentioned characteristics. Structured rules improve the act of automation support and validations of rules in business context.

5. RULE EXTRACTION MATRIX METHOD

This chapter addresses the method used in this research; the method displayed in figure 10 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 CRISPDM (KDnuggets, 2007), the method describes the steps that need to be followed to empower organizations to extract information in the form of key performance indicators and business rules. 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.

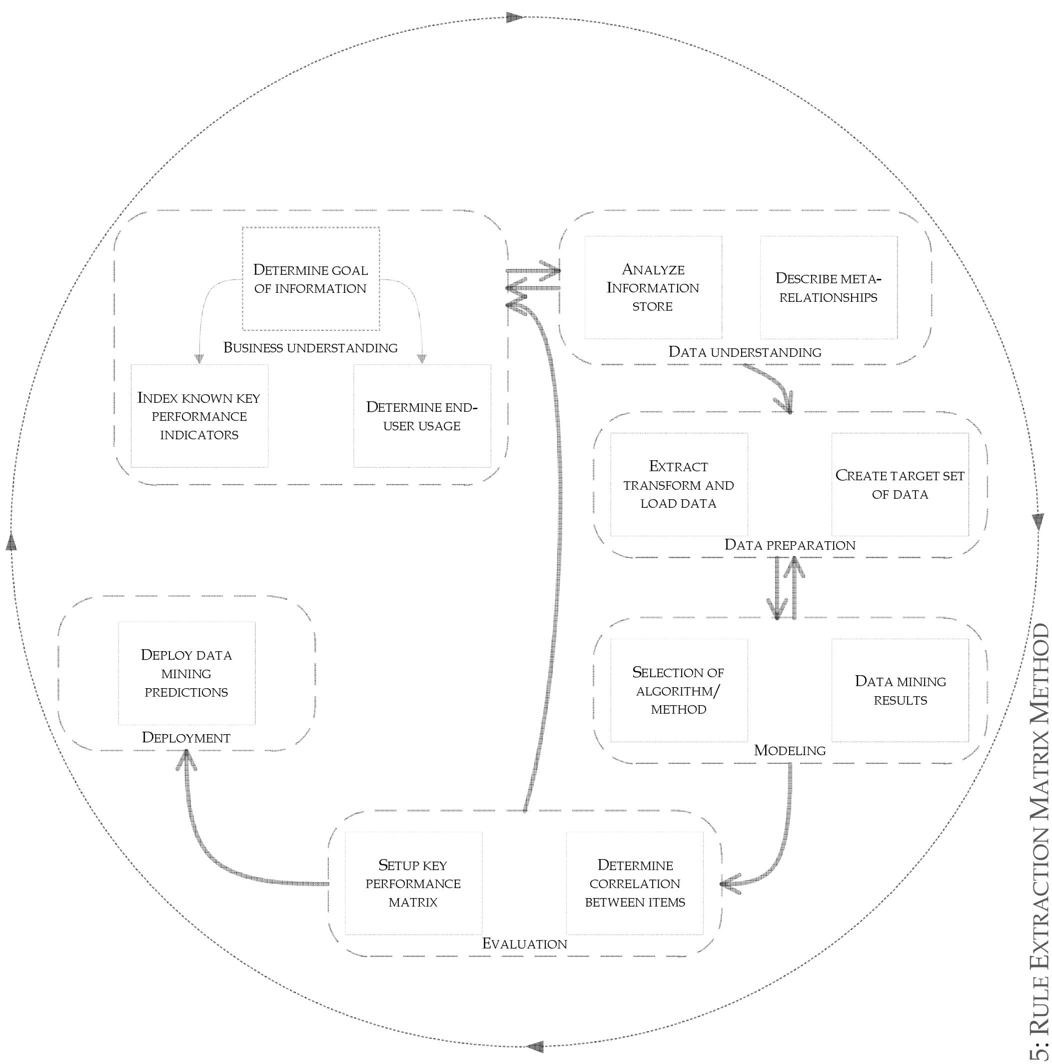

FIGURE 10: RULE EXTRACTION MATRIX (REM) METHOD

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).

TABLE 5: REM METHOD TASKS

The table above displays the phases taking place within the REM method; it consists of the steps which relate to the phases as listed in figure 10, the main preferable actors of these steps and the results following the execution of each process step.

5.1. 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, thus finding an answer on the use of the results, will they be incorporated in reports or an automated system and performing an indexing of current known key performance indicators, to assess extraction of new key performance indicators. The above results in the following steps to realize the business understanding, detailed in the figure below.

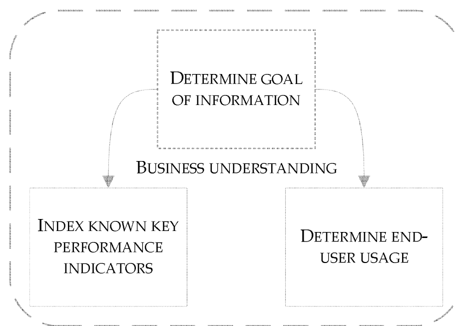

FIGURE 11: REM METHOD: BUSINESS UNDERSTANDING

The above figure shows the steps as displayed in the REM method limited towards the business understanding part of the method, the figure below shows the process deliverable diagram corresponding with the figure above.

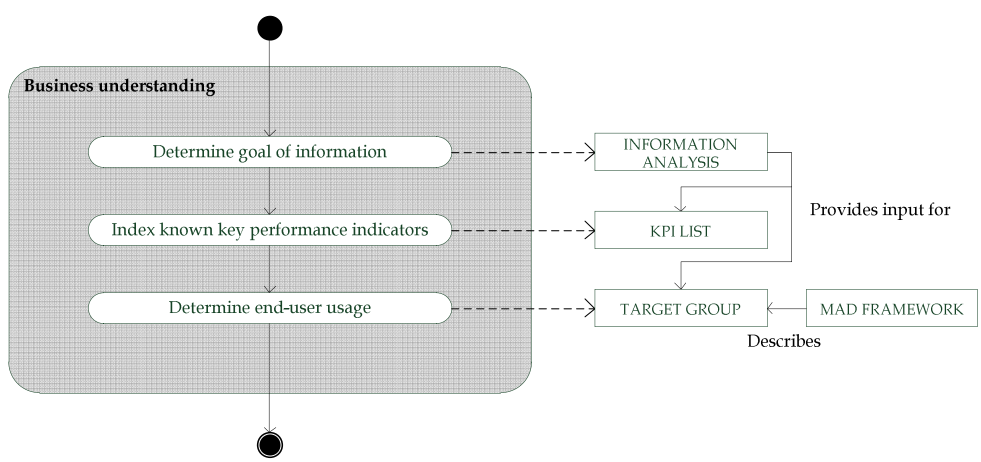

FIGURE 12: PROCESS DELIVERABLE DIAGRAM: BUSINESS UNDERSTANDING

5.1.1. Determine goal of information

The first step within the method is determining which goals are set for information extraction. This can be achieved by conversing in (un)structured interviews, analyzing previous results (continuing the circle as shown in figure 10 where the evaluation phase is linked towards the business understanding phase), as well as analyzing from a business perspective, lacking results in the form of business critical key performance indicators derived from literature/standards in the field (e.g. ISO, COPC (set of standards for Customer Service Providers) and ITIL). Figure 11 displays the main deliverable of goal determination to provide input for the next steps in the phase, knowing the indexing of known key performance indicators, as well as the end-user usage.

Figure 12 displays how the information derived from this step is bundled, in this case the information analysis which entails the basic details, such as what is expected and which information is available on the upcoming project. This step is performed by an information analyst, when looking at the MAD-framework displayed in section 4.1.3.

5.1.2. Index known key performance indicators

As shown in figure 11, the indexing of known key performance indicators relies strongly on the information goal. From this goal and the information available about the business, processes and field an organization is currently active, it can be determined which key performance indicators are currently applicable to this organization, the field or standards by which the organization operates such as mentioned in section 0. This information is gathered in a document as shown in figure 12, this list of key performance indicators is used within the evaluation phase of the method where current and analyzed key performance indicators are compared.

5.1.3. Determine end-user usage

The third step within the business understanding is to determine how the resulting business rules should be presented to the end-user. This information needs to be related to the MAD-framework as mentioned in section 4.1.3 so that the results can be adjusted to needs of the user/system in terms of presentation in terms of informal, formal or technical display of business rules and potential application development/expansion. The information gathered within this process step results in the target-group matching the categories as shown in table 3.

5.2. 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. The figure listed below displays the steps that are present in the data understanding phase.

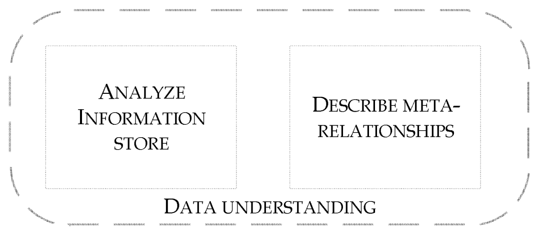

FIGURE 13: REM METHOD: DATA UNDERSTANDING

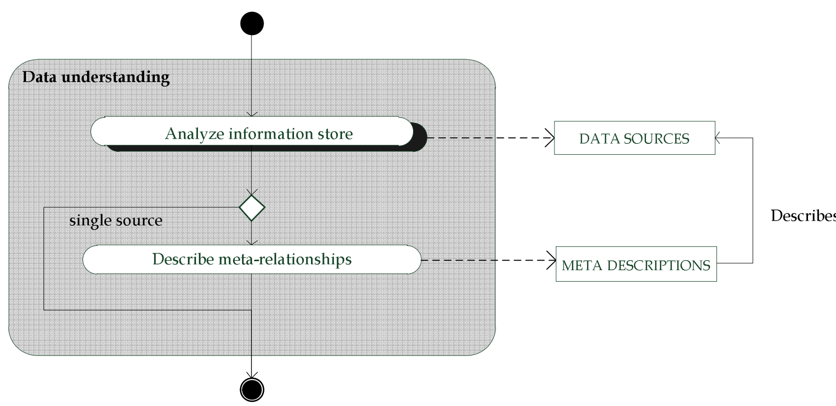

FIGURE 14: PROCESS DELIVERABLE DIAGRAM: DATA UNDERSTANDING

5.2.1. ANALYZE INFORMATION STORE

Analysis of the information store is highly dependent on the information made available in the business understanding phase. From this phase the information analysis is used to derive the needed and available information stores, resulting in an overview of data sources as described in figure 14.

5.2.2. DESCRIBE META-RELATIONSHIPS

Data sources should be described if possible with metadata and when dealing with multiple sources if applicable/possible, a description of relationships on a meta-level should be included as detailed in figure 14.

5.3. 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. The figure listed below contains the steps that occur within the data preparation phase:

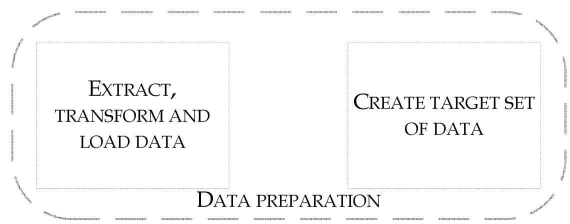

Figure 15: REM method: Data preparation

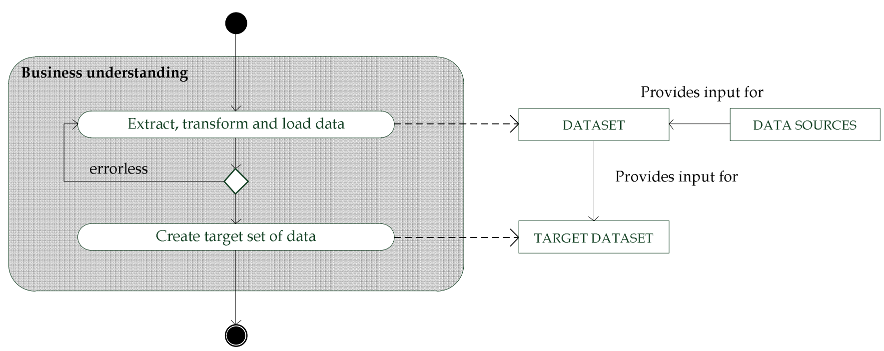

Figure 16: Process deliverable diagram: Data preparation

5.3.1. Extract, transform and load data

This step entails the creation of a clean dataset, which means the activities, such as filtering, selection and data enrichment should take place within this phase. These activities are all dependent on the information provided within the data understanding phase, the information store analysis step delivers the main details needed for the possible steps taking place in delivering a draft dataset as drawn in figure 16.

5.3.2. Create target set of data

Since processing power is still limited, though mostly financial, it is sometimes recommended to work with a target set of data based on the data quantity; the target set of data is derived from the dataset that is defined within the first step of the data preparation phase. As seen in figure 16 this step delivers the final dataset on which the modeling and data mining methods can be performed.

5.4. 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. The figure below displays the steps applicable to the modeling phase:

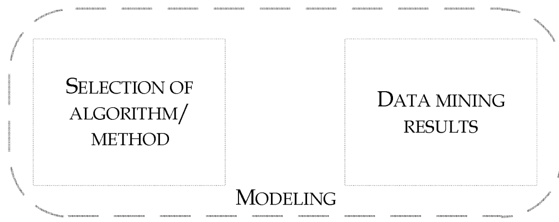

FIGURE 17: REM METHOD: MODELING

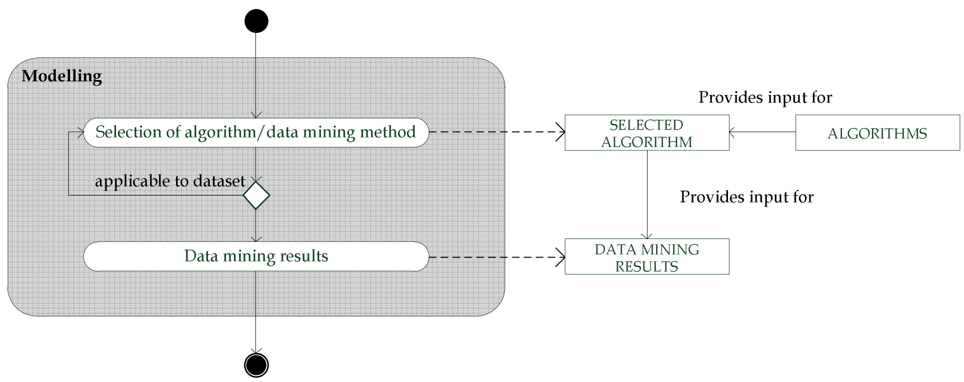

FIGURE 18: PROCESS DELIVERABLE DIAGRAM: MODELING

5.4.1. SELECTION OF ALGORITHM/METHOD

The amount of data available within the target set, as defined in the data preparation phase shown in figure 15, is of influence when the choice for a data mining algorithm/method is executed. This step results in the selection of an algorithm, which is reviewed to guarantee the possibility of applying the selection of algorithm(s) to the data.

5.4.2. DATA MINING RESULTS

As shown in figure 18, the process of obtaining the data mining results is influenced by the selected algorithm, as well as applicability towards the dataset. The results and their presentation need to match the goal of information, the given dataset and correspond to the end-users' needs.

5.5. 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. The figure below displays the steps applicable to the evaluation phase:

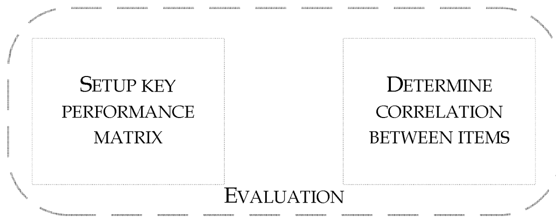

FIGURE 19: REM METHOD: EVALUATION

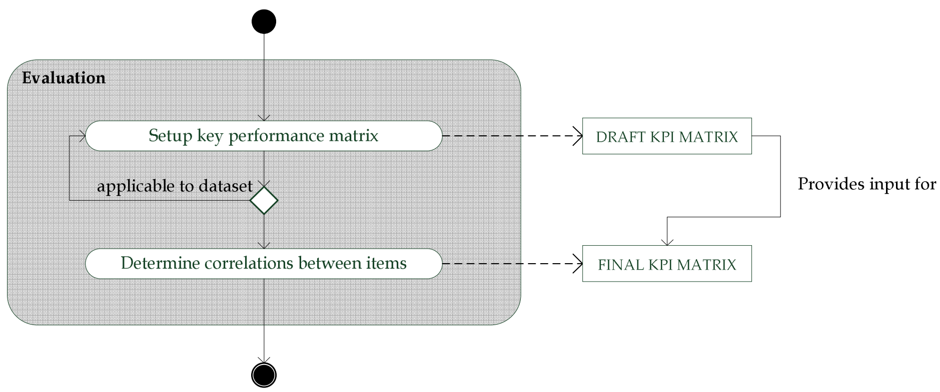

FIGURE 20: PROCESS DELIVERABLE DIAGRAM: EVALUATION

5.5.1. SETUP KEY PERFORMANCE MATRIX

The first step in the evaluation phase consists of the setup of a draft key performance matrix. This matrix consists of the preliminary results (main indicators for the business/field the organization is operating in) that are gathered in the business understanding phase in the step labeled "index known key performance indicators" and the data mining results gathered within the modeling phase. By mining the data for indicators, making use of the algorithms selected in the steps shown in figure 18, the effect of indicators on known key performance indicators can be measured.

5.5.2. DETERMINE CORRELATION BETWEEN ITEMS

The second step within the evaluation phase focuses on the correlation between the items discovered through analysis of the dataset, as well as the indicators known to the business in advance of the data mining steps of the method. By correlation of the mined items from the data mining results step, as described in section 5.4.2, and comparison of these items to the key performance indicators already known in the field, the most important items float upwards and the less important sink downwards to the list. The correlation results in the final key performance matrix which lists the key performance indicators driving performance.

5.6. 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). The figure below displays the steps applicable to the deployment phase:

Figure 21: REM method: Deployment

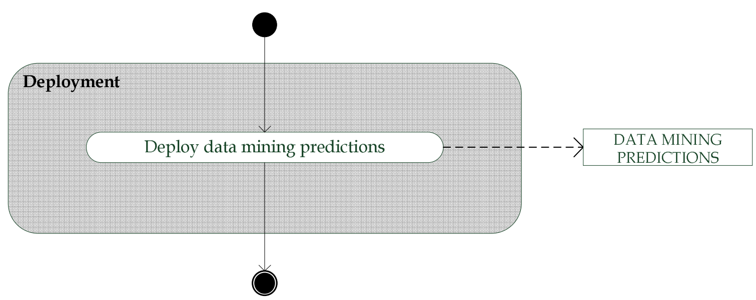

Figure 22: Process deliverable diagram: Deployment

5.6.1. Deploy data mining predictions

In order to deploy the data mining predictions, there should be an answer to the question "how will the results be used" which is presented within the business understanding phase. A possible answer to this question could be the usage of results in a report or as a structural monitored key performance indicator, which should be incorporated within a dashboard. The data mining predictions deployment consist of the traversal from key performance indicators towards key performance indicators associated with business rules.

6. CASE STUDY: OPTIMIZING PERFORMANCE AT VANAD

This chapter describes the case study performed within VANAD, using the REM method as described in detail within the previous chapter. 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.

6.1. CASE STUDY INTRODUCTION

6.1.1. CASE STUDY ENVIRONMENT

The case study is performed at VANAD B.V., which is situated in Capelle aan den IJssel. VANAD B.V. consists of approximately 1000 employees (750 F.T.E) acting worldwide at market-segments such as healthcare, transport and logistics, telecommunications and financial services. Business units incorporated within VANAD B.V. aim at servicing organizations with information and communication solutions concentrating on business process outsourcing, information technology outsourcing, managed services and telecommunications.

The research itself is situated at the telecommunications business unit servicing organizations like Achmea, Microsoft, Meeùs and SNS. The telecommunications business unit simplified business model can be drawn schematically as seen in figure 23.

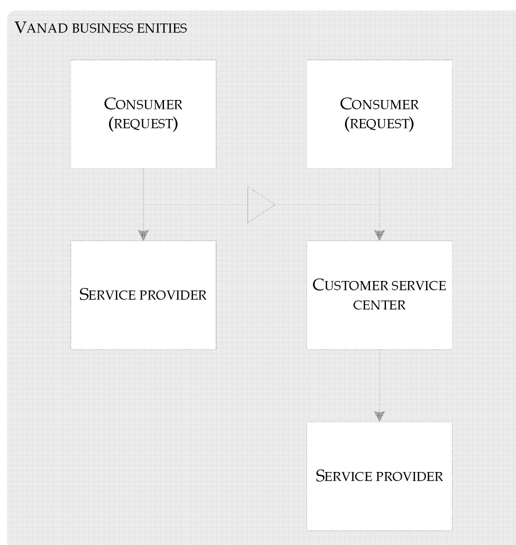

FIGURE 23: BUSINESS ENTITIES CUSTOMER SERVICE CENTRE

The figure above displays two situations; the one on the left displays immediate contact with a service provider and the one on the right shows a situation in which VANAD is acting as an intermediary for the consumer request routing it to the service provider.

The actors in the figure above are as following:

- Consumer (request) is short for the information or service an end-user requests from the service provider
- Service provider is an organization, who provides a service which end-users make use of
- Customer service centre is an organization fulfilling the further servicing of a consumer request and ultimately route the request.

Essentially, VANAD and adds one step to the process of processing a consumer request, which is handling consumer requests, formed by emails, telephone calls, text messaging and different social media or web 2.0 messages for a service centre. Customer service centers act as part of the service provider in name, as well as provide formal communication and report the outcomes to the service provider. The outcome is reported by consolidated reports or by using the service provider's technical infrastructure. By outsourcing the consumer communication channels as described above, service providers do not have to reserve office space and employees for dealing with consumer requests. One of the main advantages for dealing with these consumer requests through outsourcing is enabling the service provider to maintain a global view on its consumer movement, plus receiving detailed reports on any anomalies.

As shown in figure 23 the main actors within this case study are VANAD and the service provider in this case study further mentioned as customer of VANAD. The service provider is a large telecommunications provider active within various markets (customer and business) such as mobile, fixed and voice over IP telephony, broadband internet, interactive television and mobile Internet. Within this case study the focus lies with the customer service department concerning fixed telephony. VANAD is one of the parties to which the service provider outsourced its customer service regarding the service of fixed line and broadband internet retention. Customers of this service provider will call the Interactive Voice Response system menu and based on their dialed entries are transferred towards one of the service partners. Calls from customers to the retention desk include inquiries on cancelling their subscription, complaints on costs, (lack of) services offered by the service provider and general inquiries offered towards the customers who did not find a suitable option from the IVR-menu. VANAD's employees are trained to help the customer with regard to the service provider's standards for service and information provision.

From the previous paragraph two parties can be derived that are able to store information on customer requests; knowing the service provider, which stores the entries their customers choose, the service requests entered by VANAD's employees into their systems and VANAD which performs logging on the system side of the connection between customer and their systems (performance measurement) and quality monitoring entries. Table 6 below shows the parties involved and the systems used to store the appropriate information to successfully abide by the customer's wishes.

Parties	Owner	Data storage systems
Service provider	Manager outsourcing activities	Customer information management
VANAD	Customer contact manager (process/campaign controller)	Asterisk (performance measurement), Quality monitoring

TABLE 6: OVERVIEW PARTIES AND SYSTEMS

One of the key factors in a business-to-business relationship for organizations that actively handle consumer requests is providing customers (described as service providers in figure 23) with support. When making decisions about the customer's products and service. Support may vary from answers to questions, such as "Which changes should we make to our products to maximize consumer satisfaction?", "What do consumers or end-users think of our service and products?", "How can we maximize sales on a specific product?", "How should changes to our products be communicated" to "What are unique selling points from a consumer's perspective?"

Project managers want to take an advising role in respect to their customers; however, they do not have (enough) explicit knowledge available to provide customers with an answer to these questions as summed up in the first paragraph. Thus, project managers are not sufficiently capable of providing customers with an answer based on rationale. When questions by customers arise, project managers use average values and measurements of frequently occurring events. Part of the advice should consist of future recommendations and forecasts; however, this information is not available at present. The majority of current, available information is derived from exports, which are generated from data-sources available; mostly in the form of weekly excel exports. These generated exports offer a limited view in terms of history and details of the data.

The organization has access to three separate databases with performance data. These databases have no structure or clearly defined relations at present and all store different data, such as call times, handling times, call specifics, quality and error monitoring entries and consumer data (the latter stored on location of service providers (customer infrastructure)).

6.1.2. Case description

Currently, a large amount of data is stored in three unconnected data storage systems of which one is located at the customer's site and two systems are located at VANAD's site. This data is used for weekly reports or status updates; however, no efforts have been made to analyze how the data influences each other. An example of the before-mentioned situation is the ability of the service provider to closely monitor the outcomes of each inbound customer service request. Is the service request from a customer successful or did both parties fail to reach a solid solution to the customer service request. Both outcomes are dependent on a variety of variables such as waiting times experienced by customers, which is a direct effect of the service level as described in section 4.3 on key performance indicators, the solutions offered by VANAD's service employee and the quality of vocal techniques applied within the service request by telephone between customer and employee. Currently, the service provider and VANAD have no way to allow data interaction by key performance indicator analysis between the systems that gather customer data, asterisk data and quality monitoring information. One of VANAD's pitfalls is the current information analysis, often done 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 VANAD turns towards analyzing its current information sources. Summarizing the case leads to the following: VANAD's customer needs information on successful factors for a proposition to be accepted or in other terms for a customer inquiry to be handled successfully. By analyzing key performance indicators from a business perspective, determining their measurement by using business rules and finally providing a detailed answer towards the question "which factors are key for a successful handling of a service request?"

6.2. Business understanding

As described in the REM method, the first step is business understanding, doing so this phase incorporates the interviews, literature analysis, as well as determining the end-user usage. The steps below are numbered using the order as shown in the method.

6.2.1. Determine goal of information

The introduction and case description detail the current situation of the organization; following this the following goals for information analysis have been determined by VANAD:

i. Decisions based on information analysis.
ii. Providing the right information.

iii. Using key performance indicators and business rules to optimize performance.

Within the market VANAD operates the Customer Operation Performance Center standard offers several variables based on performance measurement and performance management.

6.2.2. Index known key performance indicators

Several known key performance indicators exist within campaigns (projects sub categorization such as retention, sales, e-mail marketing and such) executed by VANAD some of which are found within literature. The following table shows known key performance indicators in the field of inbound traffic, these key performance indicators are also found within the literature on the customer operations performance centre standard.

Common examples of key performance indicators within VANAD include service level (the amount of accepted calls within 20 seconds, plus the amount of abandoned calls within 20 seconds divided by the total calls answered, plus the total calls abandoned), average talk time (elapsed time from the moment an employee answers a phone call until disconnection of the phone call) (Cheong, Kim, & So, 2008). These common key performance indicators can be expanded by using thresholds, which are often found in service level agreements commonly constructed on the acceptance of new projects. Examples of thresholds include a service level that is at least 80% or average talking times varying from 300 seconds to 480 seconds.

Kpi	Definition and calculation
Service Level	The percentage of calls answered within a specified number of seconds (Calls answered within 20 seconds + Calls abandoned within 20 seconds)/(Total calls answered + Total calls abandoned)
Average Speed of Answer	The average time it takes for the call to be picked up by the customer service center's automatic call distribution system (Total time in queue/total number of calls handled)
Average Time in Queue	The total time in queue to reach an agent divided by the total number of calls answered (available from the automatic call distribution system)
Average Abandonment Rate	The percentage or number of customers who disconnect the call while waiting for an agent (Number of hang-ups and dropped calls/Total number of calls offered)
Percentage of Calls Closed on First Contact	The percentage of calls that properly and completely address the customer's needs in the first call (calls closed on the first contact)/(total number of calls closed by all agents)
Adherence to Schedule	The average time the agent is in their seat as per the work force management schedule
Average Talk Time	The elapsed time from when an agent answers a call until the agent disconnects (available from the automatic call distribution system)
Average After Call Work Time	The time an agent spends completing a transaction precipitated by a phone call after the call is released
Employment Turnover Rate	The percentage of agents who leave the customer service center to work elsewhere

Percentage of Calls Blocked	The percentage of calls that received the busy (engaged) tone (Total number of calls offered by the network - Total number of calls processed by switch)/(Total number of calls offered by network)
Customer Satisfaction	The customers' subjective views as derived from questionnaires average score or rating of sampled calls based on third party customer survey

TABLE 7: KEY PERFORMANCE INDICATORS APPLICABLE TO INBOUND TRAFFIC(CHEONG, KIM, & SO, 2008)

Most of the above metrics are monitored by the systems of VANAD except for the ones marked in grey. These are the ones monitored by the service provider's systems and not shared with VANAD and the indicators not measured by VANAD since the data is not formatted or collected.

Next to the Key performance indicators listed above, VANAD defined a quality monitoring system containing call grading on a five point Likert-scale, calls are graded using thirteen different aspects of calls such as summarization of a resolution, administrative handling. The items monitored are weighed and result in a summarization for the whole campaign on a weekly basis.

The service provider stores the customer data and details reports towards VANAD on how their performance is compared to other outsourcing parties on a conversion level (accepted propositions versus denied propositions), as well as other performance metrics.

6.2.3. DETERMINE END-USER USAGE

The end-user usage is two-fold, firstly, the organization would like to use the results in their automated systems, secondly, the results should also be interpretable by project managers in order to support their decision making processes. When looking at the roles within the organization we can see that most of the employees working with operational reports rate as analysts or workers in the MAD framework as displayed table 3.

The organization tends to make use of the results by dynamic reports and operational queries. Using the results from the final key performance matrix, decisions can be supported using business rules. The current situation offers room for implementing business rules in the systems available within VANAD, as well as the opportunity to provide business rules to the customer contact managers to aid in their decision making process.

The form of presentation should be technical/formal to allow interpretation by the systems and informal for use by operational managers.

6.3. DATA UNDERSTANDING

The second step within the REM method is data understanding; the goal of this phase is analysis of the data-stores to gain an understanding of the limits of the data and a feel for the possibilities. The analysis of data results in an overview of information stores and metadata.

6.3.1. Analyze information store

The following data-stores are available:

The data-store located with the service provider storing information about project results such as "conversie" ratings (the amount of accepted propositions/solutions) and service-levels on the inbound campaign. The data-store located with the service provider embodies results of multiple campaigns, as well as the results of multiple service-centers, unfortunately due to security reasons and the fact that this data is sensitive a comparison between service-centers is not possible.

Quality monitoring located with VANAD storing information on the service requests that take place within several campaigns within VANAD. The properties monitored here are key components of a service request such as the right solution offered and careful analysis taken place within a service request.

Asterisk (communications server software) located in VANAD, storing specific details also called sip variables such as incoming lines, the handling times of service requests, service levels about service requests, Asterisk provides the data of which the main part of the variables mentioned in table 7 can be calculated.

6.3.2. Describe meta-relationships

The customer information store has no specific relations to the others than the date a service request by telephone took place and the number from which was called, this way only customers who do not shield their telephone number are analyzed. Both databases stored with VANAD are connected by the use of unique ID's. The databases are used for more than the above mentioned views; however, to create a working environment and keep the data as clean as possible the above views/exports are used to determine the correlation between variables.

An overview of information available within the data-stores is provided in Appendix B: Data structure which details the data and meta-data for each data store. Figure 24 shows an overview of the data stores and their locations. The difficulty exists in the measuring levels, the way the service provider measures performance differs from the way VANAD does. VANAD takes a (random) selection of service requests held for the service provider and tries to analyze these service requests on a fixed set of properties, thus resulting in a sample. The service provider focuses on variables that can be measured by calculations on the whole dataset, such as service levels and the percentage of accepted solutions set out against all calls (conversion-percentage). To address these issues the measuring (scaling) level, which compares the service provider's performance measurement to the performance measurement executed by VANAD, needs to be adjusted towards a weekly scale. Comparison of quality and sip (both stored at VANAD) variables can be performed at more detailed scales since both measurements are based on a service request-ID.

The resulting dataset is shown in table 8 containing a metadata overview. 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 optional. As shown in section 10.2.1. a 0..* relation exists between the asterisk_cim and quality_monitoring.

The dataset that combines the three data-stores available results in a data-structure containing the variables measured, brought back to the weekly measuring level of the service provider since it is possible to transform the analysis performed by VANAD to a weekly level.

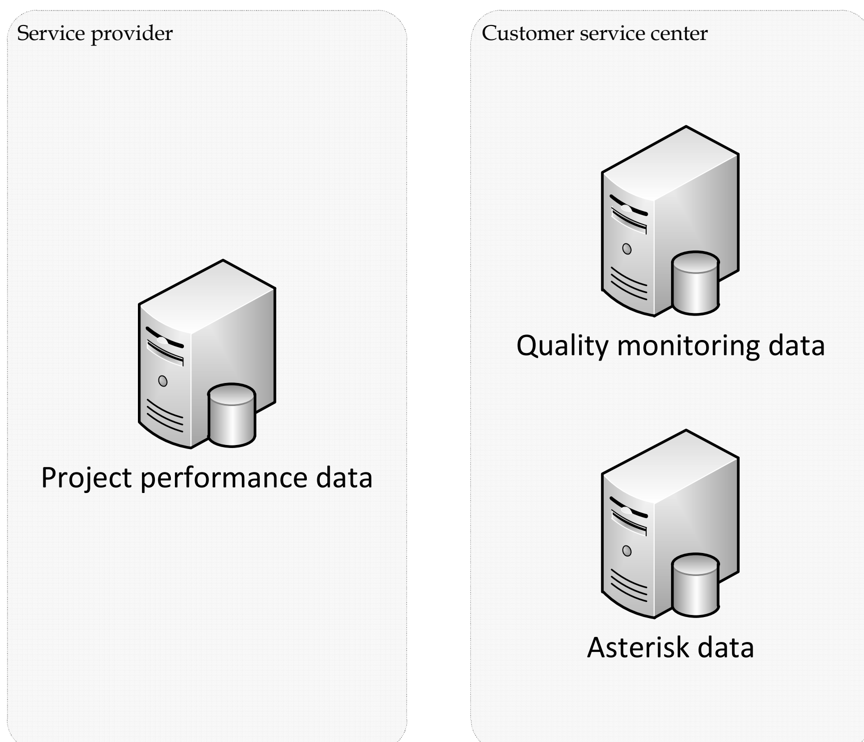

FIGURE 24: DATA STORES AND LOCATIONS

Role	Name	Type	Statistics	Range
ID	asterisk_uid	real	count = 3971 records	-
regular	date_received	date_time	length = 451 days	-
regular	strategy_type	binominal	mode = queue_based (2281)	-
regular	strategy	polynominal	mode = queues_liar (2281)	-
regular	call_time(minutes)	real	-	-
regular	wait(boolean)	binominal	-	-
regular	waittime(minutes)	real	-	-
regular	transfered	binominal	-	-
regular	Administratieve afhandeling	integer	-	[0.000 ; 4.000]
regular	Afsluiting	integer	-	[0.000 ; 5.000]
regular	Behoefte(n)bepaling	integer	-	[0.000 ; 4.000]
regular	Gespreksvoering	integer	-	[0.000 ; 4.000]
regular	Introductie	integer	-	[0.000 ; 5.000]
regular	Klanttevredenheid	integer	-	[0.000 ; 5.000]
regular	Leiding behouden	integer	-	[0.000 ; 4.000]
regular	Luisteren	integer	-	[0.000 ; 4.000]
regular	Productkennis	integer	-	[0.000 ; 5.000]
regular	Propositie(s) en Analyse	integer	-	[0.000 ; 4.000]
regular	Samenvatting	integer	-	[0.000 ; 4.000]
regular	Structuur	integer	-	[0.000 ; 4.000]
regular	Vragen	integer	-	[0.000 ; 4.000]
regular	Conversie ADSL nieuw	real	-	-
regular	Conversie IPB nieuw	real	-	-
regular	Conversie BB-nieuw totaal	real	-	-
regular	Conversie ADSL bestaand	real	-	-
regular	Conversie IPB-bestaand	real	-	-
regular	Conversie BB-bestaand totaal	real	-	-
regular	Conversie ADSL totaal	real	-	-
regular	Conversie IPB-totaal	real	-	-
regular	Conversie BB nieuw + bestaand totaal	real	-	-
regular	Conversie ITV nieuw + migr.	real	-	-
regular	Conversie Digitenne nieuw + migr.	integer	-	-
regular	Conversie TV Totaal	real	-	-
regular	Conversie VAS totaal	real	-	-
regular	Conversie Totaal (BB-totaal + TV totaal + VAS)	real	-	-
regular	Service-level	real	-	-

TABLE 8: METADATA RESULTING DATA

6.3.2.1. Trigger hypothesis

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 the systems on the long run and thus tries to find specific relations between performance measured by the service provider and performance measured by VANAD.

6.4. 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.

6.4.1. Extract, transform and load data

The datasets in question are exports (thus extracts) from each mentioned database, knowing customer information management, asterisk and quality monitoring over several longer periods to allow for comparison and validation. Essentially, data transformation cleanses the data of malformed entries, which are not processed by the data mining tool, this is performed to prevent malformed entries. 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.

6.4.2. Create target set of data

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 loaded into the data mining tool and finalized for algorithmic analysis, thus analyzing errors and providing the data mining tool with the right fields and meta-settings for those fields.

6.5. Modeling phase

The fourth step of the REM method is the modeling phase, resulting in the selection of an algorithm suitable for the data mining needs of the organization, as well as the execution of said/selected algorithm.

6.5.1. Selection of algorithm/method

When looking at the first phase of this case study, the organization wants to know which factors are key for a successful service request. Based on the trigger hypothesis, defined in section 6.3.2.1, 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. From descriptive methods as described in section 4.2 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. The process of binning the variables is performed by the data mining tool. 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 as described in section 6.2.2. 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. The data mining tool allows a decision tree to be transformed in to a set of business rules b by internal application of association rules to the same dataset as used when defining the decision trees.

6.5.2. Data mining results

The results are elaborated on using the Repository overview as a starting point, Rapid Miner allows to store processes, as well as results in its repository, thus enabling the steps from data cleansing to results to be recorded and archived. Figure 25 shows the Rapid Miner repository containing the dataset, processes and results available for this research. The dataset has been cleansed using the standard ETL-tools available within Rapid Miner, resulting in the dataset that can be found in Appendix B: Data structure describing the meta-data for the resulting data structure.

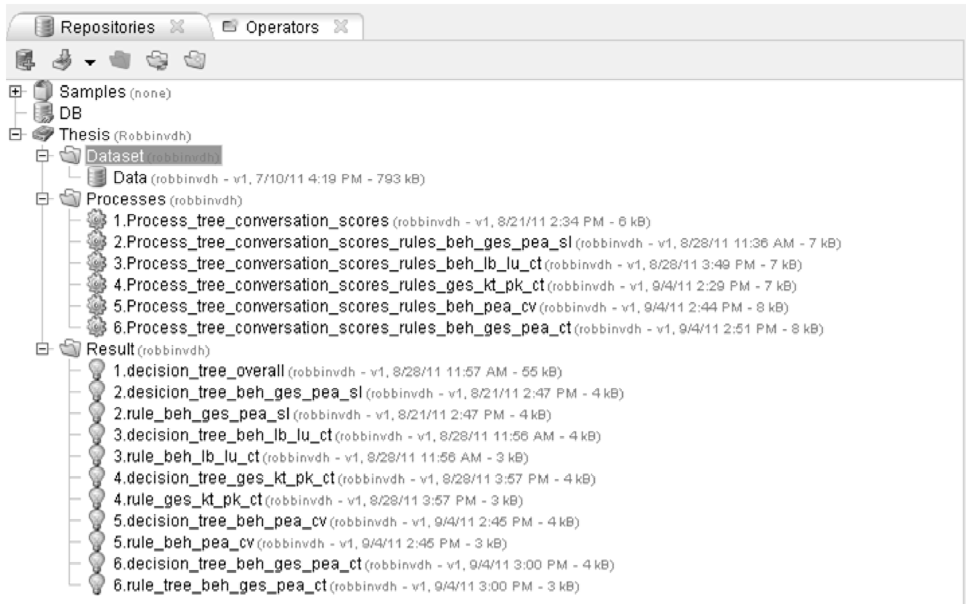

FIGURE 25: RAPID MINER REPOSITORY OVERVIEW

The first step in gathering results was to obtain an overview of the data and how variables influence each other, in order to achieve this a process has to be created using Rapid Miner. The creation of the process and the functions of the operators available for Rapid Miner are not dealt with in the description of the results. The process is shown in the figure below.

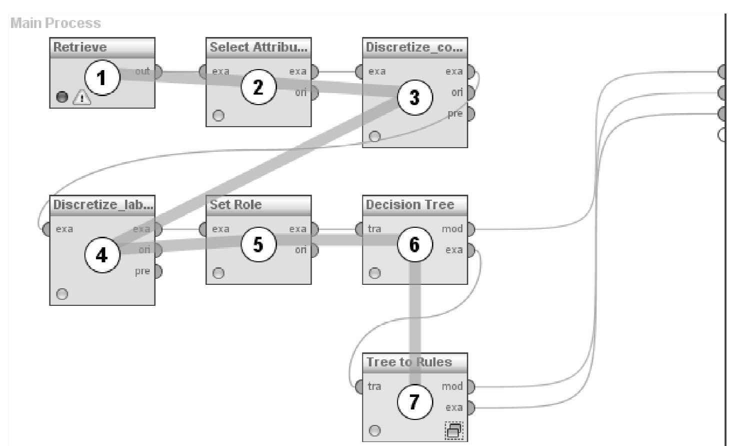

FIGURE 26: PROCESS OVERVIEW RAPIDMINER

The figure above shows the six steps which include retrieving the dataset from the repository, selecting the attributes (variables), necessary to obtaining the results, discretizing (categorization of values) the variables, as well as the target (label) variable, setting the role of the dependent variable and creating the decision tree.

Running the process results in the following:

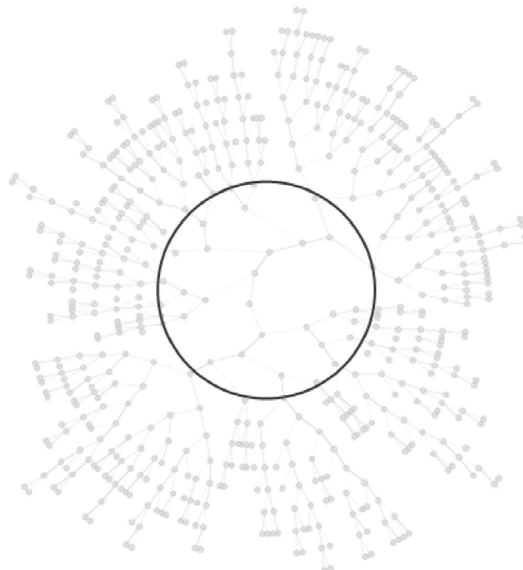

FIGURE 27: DECISION TREE

Zooming in on the above decision tree (the area marked by the grey circle) results in the following image:

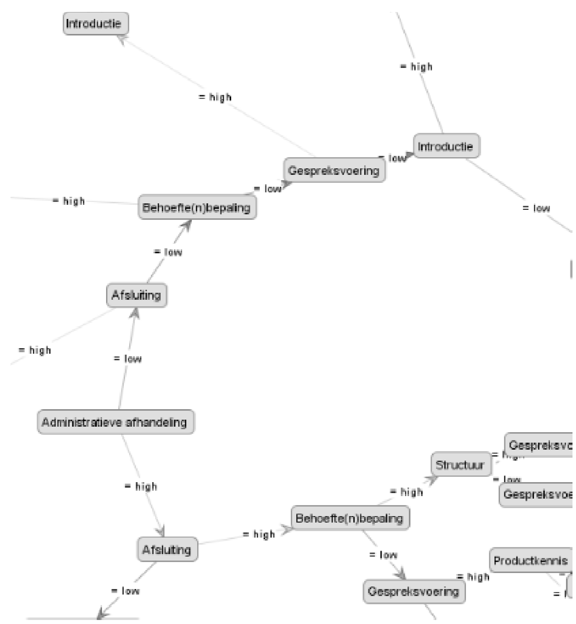

FIGURE 28: DECISION TREE ZOOMED

Using the result-overview as detailed in figure 27 and figure 28, the narrowing down of the results is performed by defining new processes, which select certain attributes and form these into rules that define the influence of variables on target (dependent) variables. This step in result gathering is essentially, taking a broader overview and narrow down the results using the software at hand.

The part of narrowing down the results meant focusing on the (dependent) variables which have the biggest business impact (service-level and conversie) and selecting variables that are of influence on the dependent variables. The dependent variables have been selected from a short-list of variables that VANAD needs to influence, in order to perform at the level defined by contractual agreements. The final resulting short-list is constructed by the top three priority variables (service-level, conversie, call time).

6.5.2.1. Description (current situation)

The following shows the results obtained by narrowing down and zooming in on the strongest relations within figure 27. The section-title shows the variables selected and the dependent variable marked by brackets (dependent). Using the process overview as shown in figure 26, the data is analyzed and a construction of the resulting decision tree is shown in each sub-section. Firstly, the dataset is selected as shown in step one and fed to the selection of attributes as displayed in step two. The attributes which are selected are named in each section title, for example, in the first section below, these are behoeftebepaling, gespreksvoering, propositie en analyse and service-level (dependent/target). Creating a decision tree in Rapid Miner requires the use of non-numerical attributes and a target variable. To construct non numerical attributes the attributes are binned (discretized) using step three and step four as shown in figure 26, step four is in preparation of setting the target variable, in the section below this is the dependent variable service level. Step four results in a binned target variable (table 9 shows the thresholds used for the binning process of the target varbiables), plus the variables are ready to be fed to step five, in which the target variable is selected. Now the attributes are prepared and the label variable has been set, the decision tree can be constructed. This is performed in step six of the process overview, resulting in the decision trees as displayed in each section.

Target	Binned
Service level	Below (<65%), Above (>65%)
Call time	Low (<10), High (>10)
Conversie	Below (<15%), Above (>15%)

Table 9: Overview binning (categorization) target variables

6.5.2.1.1. Behoeftebepaling, gespreksvoering, propositie en analyse and service level (dependent)

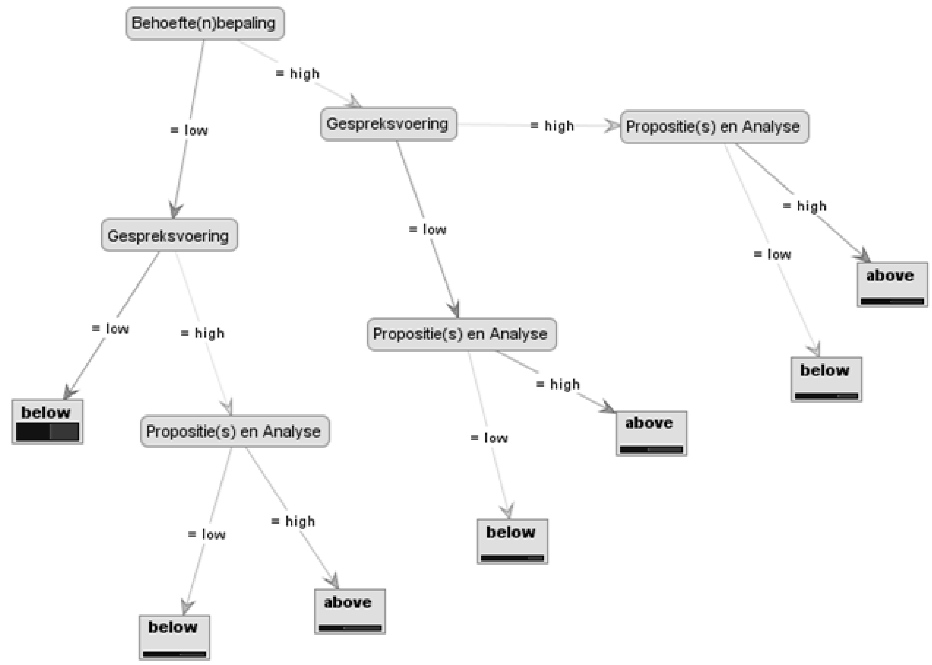

Figure 29: Decision tree supporting section 6.5.2.1.1

Figure 29 shows the variables that have the biggest influence on the service-level variable (The percentage of calls answered within a specified number of seconds (Calls answered within 20 seconds + Calls abandoned within 20 seconds)/(Total calls answered + Total calls abandoned)) from analyzing the results and applying discretization, it was found 4 of the five point likert-scale levels were actually used, thus discretization of the quality variables was performed by categorizing each value <=2 as low and the variables >2 as high this applies to the rest of the results as well. The categorization of the service level variable was performed by taking the specified minimal service level as found in the contractual agreements.

Preliminary conclusions (found by analyzing the above results) are that a positive increase in the selected quality results, which leads to an increase in measured service levels.

6.5.2.1.2. Behoeftebepaling, leiding behouden, luisteren and call time (dependent)

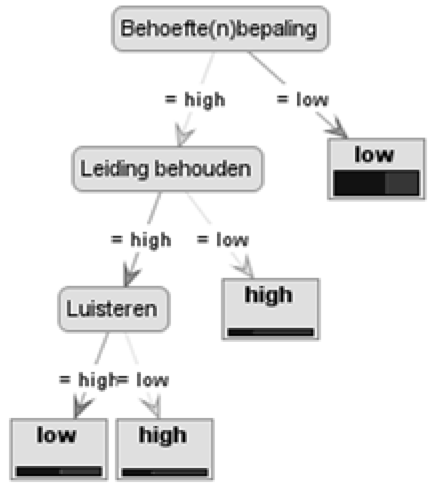

Figure 30: Decision tree supporting section 6.5.2.1.2

Figure 30 shows the variables influence call times (the time it takes for a service request to be finished, including administration). The above figure shows that a positive approach to the customer results in calltimes below maximum (10 minutes), from the figure above it is found that variables have multiple results, for example, "behoefte(n)bepaling" has two possible outcomes, if it is low the expected calltimes are also low; however, when combined with a positive "leiding behouden" and "luisteren" the resulting call times are also low.

Preliminary conclusions found from analysis of the above results show that a positive score on the above quality aspects also influence the call time in a positive manner.

6.5.2.1.3. Gespreksvoering, klanttevredenheid, productkennis and call time (dependent)

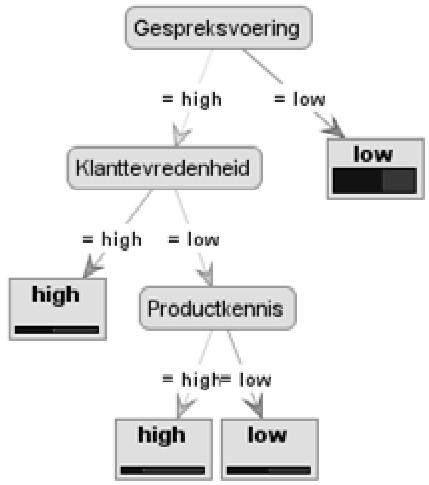

Figure 31: Decision tree supporting section 6.5.2.1.3

The above figure shows the influence of "gespreksvoering", "klanttevredenheid" and "productkennis" on the dependent variable call time, from this it can be derived that a decrease in quality aspects results in lower call times, this is expected since the above quality aspects relate to the analysis of problems and customer situations, thus taking time to perform a thorough analysis.

Preliminary conclusions found from analyzing the above results and the results from the previous section show that call times can be influenced in a positive and negative manner, when looking at quality aspects.

6.5.2.1.4. Behoeftebepaling, gespreksvoering, propositie en analyse and call time (Dependent)

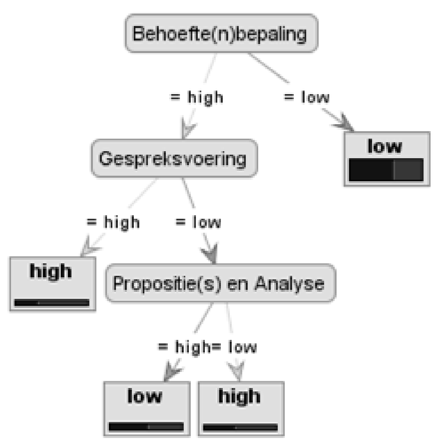

Figure 32: Decision tree supporting section 6.5.2.1.4

Figure 32 shows the influence of the quality variables from section 6.5.2.1.4 on the dependent variable call time. Negative quality aspects result in positive call times, except for the quality variable “propositie en analyse” if this is regarded in combination with a low on “gespreksvoering” plus a high on “behoeftebepaling” the resulting call time is regarded as low.

Preliminary conclusions support based on the above support the drawn conclusion from the previous section, call times can be positively and negatively influenced by quality aspects.

6.5.2.1.5. Behoeftebepaling, propositie en analyse and conversie totaal (Dependent)

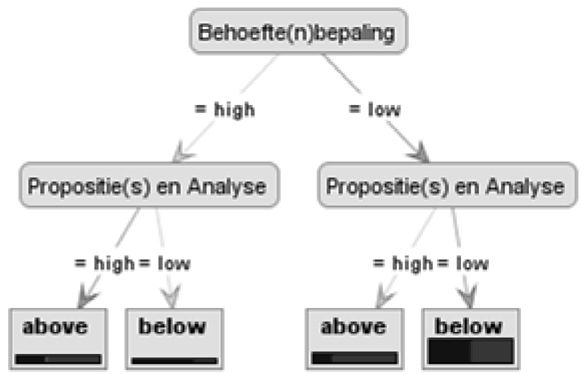

Figure 33: Decision tree supporting section 6.5.2.1.5

The above figure shows the influence of the quality variables “behoeftebepaling” and “propositie en analyse” on the variable “conversie (totaal)”.

The main conclusion found by analyzing the above results is that “conversie (totaal)” (the amount of propositions (solutions/new products) accepted by the customer) is positively influenced by “behoeftebepaling” and “propositie en analyse”, when both are above the threshold for categorization, as high or low the conversion is positively influenced; however, from the above it can also be derived that the influence of “propositie en analyse” is greater than that of “behoeftebepaling” since a low “behoeftebepaling” and a high “propositie en analyse” still results in a conversion above target.

6.5.2.2. Prediction (rules)

This section summarizes the detailed business rules that were extracted from the data and tree diagrams that are listed in the previous section(s). Each tree section is labeled as a rule and is created by using the results acquired in step six as shown in figure 26 in step seven as the traversal from a decision tree to a business rule. The format shows the business rules that have been derived from the decision trees and shows the amount of cases in brackets for which the business rule is applicable (first number) and the amount of cases excluded from the business rule (second number), exclusion can occur on non applicability or no available data due to missing variable entries.

6.5.2.2.1. Behoeftebepaling, gespreksvoering, propositie en analyse and service level (dependent)

- if Behoefte(n)bepaling = high and Gespreksvoering = high and Propositie(s) en Analyse = high then above (33 / 34)
- if Behoefte(n)bepaling = high and Gespreksvoering = high and Propositie(s) en Analyse = low then below (7 / 3)
- if Behoefte(n)bepaling = high and Gespreksvoering = low and Propositie(s) en Analyse = high then above (191 / 224)
- if Behoefte(n)bepaling = high and Gespreksvoering = low and Propositie(s) en Analyse = low then below (77 / 24)
- if Behoefte(n)bepaling = low and Gespreksvoering = high and Propositie(s) en Analyse = high then above (22 / 31)
- if Behoefte(n)bepaling = low and Gespreksvoering = high and Propositie(s) en Analyse = low then below (29 / 19)
- if Behoefte(n)bepaling = low and Gespreksvoering = low then below (1834 / 1443)

6.5.2.2.2. Behoeftebepaling, leiding behouden, luisteren and call time (dependent)

- if Behoefte(n)bepaling = high and Leiding behouden = high and Luisteren = high then low (275 / 244)
- if Behoefte(n)bepaling = high and Leiding behouden = high and Luisteren = low then high (8 / 14)
- if Behoefte(n)bepaling = high and Leiding behouden = low then high (16 / 36)
- if Behoefte(n)bepaling = low then low (2101 / 1277)

6.5.2.2.3. Gespreksvoering, klanttevredenheid, productkennis and call time (dependent)

- if Gespreksvoering = high and Klanttevredenheid = high then high (61 / 69)
- if Gespreksvoering = high and Klanttevredenheid = low and Productkennis = high then high (3 / 8)

- if Gespreksvoering = high and Klanttevredenheid = low and Productkennis = low then low (19 / 18)
- if Gespreksvoering = low then low (2317 / 1476)

6.5.2.2.4. Behoeftebepaling, gespreksvoering, propositie en analyse and call time (dependent)

- if Behoefte(n)bepaling = high and Gespreksvoering = high then high (25 / 52)
- if Behoefte(n)bepaling = high and Gespreksvoering = low and Propositie(s) en Analyse = high then low (229 / 186)
- if Behoefte(n)bepaling = high and Gespreksvoering = low and Propositie(s) en Analyse = low then high (45 / 56)
- if Behoefte(n)bepaling = low then low (2101 / 1277)

6.5.2.2.5. Behoeftebepaling, propositie en analyse and conversie totaal (dependent)

- if Behoefte(n)bepaling = high and Propositie(s) en Analyse = high then above (177 / 305)
- if Behoefte(n)bepaling = high and Propositie(s) en Analyse = low then below (82 / 29)
- if Behoefte(n)bepaling = low and Propositie(s) en Analyse = high then above (202 / 581)
- if Behoefte(n)bepaling = low and Propositie(s) en Analyse = low then below (1353 / 1242)

6.6. Evaluation phase

The fifth step in the REM method is the evaluation phase, by the setup of a matrix that ultimately results in an overview of correlated weighted items key performance indicators and business rules can be derived from the matrix.

6.6.1. Setup key performance matrix

The key performance matrix essentially shows which variables influence the dependent variables and which effects an increase in independent variables has on the measured dependent variables ("service-level", "call time", "conversie(totaal)"). It should be noted when looking at this matrix that an increase in the dependent variables "service-level" and "conversie(totaal)" is desired, the dependent variable "call-time" should decrease when looking at operational results.

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)
Leiding behouden	Call-time	Positive (Decrease)
Luisteren	Call-time	Positive (Decrease)
Gespreksvoering	Call-time	Negative (Increase)
Klanttevredenheid	Call-time	Negative (Increase)
Productkennis	Call-time	Negative (Increase)
Propositie en Analyse	Call-time	Positive (Decrease)
Behoeftebepaling	Conversie (totaal)	Positive (Increase)
Propositie en Analyse	Conversie (totaal)	Positive (Increase)

Table 10: Draft key performance matrix

The above matrix shows (on measured dependent variables) the effects that independent variables have. Using this matrix, the customer service provider will be able to direct its focus by using a combination of the effects measured above. For example when launching a campaign that focuses most on availability of agents (for example emergency/standby services for customers) focus should be put on service-levels and call-times (handling times). In order for the campaign to achieve its targets, the quality variables "behoeftebepaling", "gespreksvoering", "propositie en analyse", "leiding behouden", "luisteren" and "klanttevredenheid" should be held up to a high standard.

6.6.2. Determine correlation between items

The draft key performance matrix shows the 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. The following figure shows the correlations between the dependent variables and the 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 34: Correlations indicators and dependent 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. Three variables stand out in terms of effects on the three dependent variables (“service-level”, “call-time” and “conversie(totaal)”) with a correlation coefficient of at least (-).475 (r^2 * 100 = 22,56 % variation of the dependent variable is explained by the independent variable) the quality variables “behoeftebepaling”, “propositie en analyse” and “gespreksvoering” have the largest effects and show the biggest influences. The resulting KPI matrix, as shown in the table below, lists the effects.

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)

Table 11: Final key performance matrix

One final note on the final key performance matrix is the (direct) influence the call time exerts on the service-level, a general assumption made by project managers. By analyzing of the call times and the resulting service-level, it can be seen that the service-level is influenced by more variables than just the call time. One example of this is the capacity of the employees, if this is exceeded (more service calls than forecasted) the service-level will not be met. The service-level is highly dependent on a low abandonment rate and high volume of answered calls, if one of these is influenced by an positive or negative effect the impact on the service level will be high.

6.7. Deployment phase

6.7.1. Deploy data mining predictions

The data mining predictions that are shown in section 6.5.2.2 allow integration with the systems that VANAD uses to monitor performance at the individual level. Using these rules allows integration with systems used at VANAD that help with SWOT-analysis, essentially applying the analyze-functionality as shown in table 3. A description of such a case would be the need for training in the service request quality aspects “gespreksvoering” and “propositie en analyse” to positively influence the dependent variables “conversie” and “call-time”.

Next to the implications towards the systems these results have, the draft key performance matrix, as well as 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. A real life question from a customer could be “How do we guarantee service-levels while upholding our current project results?“ Through prioritizing the key performance indicators and the independent performance indicators, this question now has a measurable answer, in terms of which performance indicators should be measured and should receive the most focus. The three independent variables shown in the final key performance matrix are influencing the various key performance indicators. As one can conclude, maintaining project results in terms of low call times can influence the ability to uphold to the service level and conversion. For example, maintaining a high “behoeftebepaling” results in high call times (not desired), though this also results in maintaining the service level and conversions at the desired level.

7. Evaluation

This chapter details the evaluation of the results acquired through the REM method, the evaluation is performed by structured interviews and a presentation of the acquired results in the form of the final key performance matrix towards the respondents. The interview questions are listed in Appendix C: interview questions. The respondents for this interview are categorized as executives/managers/analysts as per the categorization by the MAD framework.

7.1. Interviews

In order to present the results to the respondents, the form of business rules as shown in section 6.5.2.2 needs to be addressed. These business rules are currently in technical form, by combining these results and key performance thresholds. The combination of these results is performed by using the thresholds set when binning the variables within the data mining application. For example, service-level is binned in to two categories knowing: high (above 65%) and low (below 65%) the same has been performed for the independent variables which also have been binned. Combining the business rules and the parameters used for binning the variables results in the following describing business rules:

- In order to keep the service level at a minimum of 65% “behoeftebepaling” should be at least 2 on a scale of 5.
- In order to keep the service level at a minimum of 65% “gespreksvoering” should be at least 2 on a scale of 5.
- In order to keep the service level at a minimum of 65% “propositie en analyse” should be at least 2 on a scale of 5.
- In order to keep call times at a maximum of 10 minutes “behoeftebepaling” should be at most 2 on a scale of 5.
- In order to keep call times at a maximum of 10 minutes “gespreskvoering” should be at most 2 on a scale of 5.
- In order to keep call times at a maximum of 10 minutes “propositie en analyse” should be at least 2 on a scale of 5.
- In order to keep conversie at a minimum of 15% “behoeftebepaling” should be at least 2 on a scale of 5.
- In order to keep conversie at a minimum of 15% “propositie en analyse” should be at least 2 on a scale of 5.

Reviewing the above, it should be noted that while call times directly influence the service-level (the less time a call takes the higher amount of calls an employee can perform), the service-level is highly influenced by handling times (short for time it takes to register the calls in the systems of the service provider) and the amount of calls received per interval, if the capacity of employees is exceeded the service-level will not be met.

Analysis on the possible relation between service level and call times is performed using the same process as detailed in figure 26; however, this time the service level was set as the dependent variable and call times were set as the independent (binned) variable. By analysis, it was found that even if call times are low (less than 10 minutes), the service-level often results in below the target of 65%.

- if call_time(minutes) = high then below (848 / 723)
- if call_time(minutes) = low then below (1345 / 1055)

The above is the rule result of displaying the relationship between call times (above or below 10 minutes) and a service-level of 65%.

It is also possible that by a low performance (quality-wise), a customer decides to repeat their request to the customer service provider, thus increasing the amount of calls received and directly influencing the variables that influence the service-level. Thus "behoeftebepaling" can both have a positive, as well as a negative influence on time-based variables.

7.2. Interview findings

This section summarizes the interview response provided by the managers, which are presented with the interview questions; Appendix D: interview response details the exact answers provided by the respondents on the interview questions. Table 12 shows an overview of the respondents that have been approached for an interview.

	Respondent one	Respondent two
Function	Medior manager	Senior Analyst/Manager
Affiliation	Internal (project)	External
Expertise	Project control, budgeting and performance management.	Data analysis, pattern visualization.

Table 12: Overview of respondents

Converging strategic decisions into tacit and operational control

The respondents one and two feel that currently communication concerning quality variables and indicators is performed at the strategic level and there is no well defined process or procedure available in terms of converging the decisions top down to the operational level, respondent one feels the results that the REM-method delivers allow for a solid conversion of strategic to tactical and operational levels.

Managing projects by applying timely notification and response times

Respondents one and two describe the decision making processes a customer enrolls and describe the need for timely notification, plus strategic decision adjustment is essential to manage a project well, as well as provide useful information to the service providers. Full transparency is also a keyword in this, transparency allows to draw conclusions based on facts in support of this conclusion. Respondent one states the REM-method lends itself to provide decision supporting information to the service providers, in terms of performance indicators.

Using key performance indicators to reach targets and perform at the maximum

Respondents one and two detail the relationships which are currently drawn based on tacit knowledge. The relationships drawn by the REM-method show similarities, as well as differences such as: call-time influencing the service-level directly. The respondents feel the results gathered have motivational influences, as well as decision supporting characteristics; however, for these properties to fully maximize their potential service provider support is needed. This is only achieved by full transparency.

Applying relations to multiple campaigns and graphical overview

Respondents one and two feel the decision trees presented are a useful addition to the current key performance overviews, since they offer insights in the relationships between the key performance indicators. In order to communicate results towards the customer these relationships have to be more graphical and supported by (historical) numerical graphics. Higher accuracy is achieved by applying the business rules found to multiple projects, plus verify if the relations found are also applicable to other campaigns.

Driving performance by transparency and commitment

Summarizing the above the responses show in order to reach maximum performance, the service providers should be allowed full transparency to the results gathered, next to this commitment from the service provider should exist towards reaching a predefined quality level and resulting budgets should be made available. Translation to the operational and tactical levels should occur by repetition and selective communication of results. Finally, the most important factor is the communication of results to the executing employee; these employees should be made aware of the impacts on results certain actions they consciously or unconsciously make.

7.3. Deliverables

This section discusses the results gathered through the case study and the appliance of the REM method.

7.3.1. Rule Extraction Matrix method

The steps of the REM method have been applied to the case study at VANAD, by the appliance of REM method new business rules were extracted from the data at hand. The extensions applied to the CRISPDM model allowed for successful application to the case study. The extensions fitted to the CRISPDM model allowed to obtain 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.

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.

7.3.2. Performance matrix evaluation

The performance matrix, as a result of the case study, is the main deliverable for this research. Given the data and the research path chosen; this is the direct result of the analysis that took place. Stated by the input provided and the analysis performed by the data mining tool, these decision trees can be replicated using the same data. The performance matrix lends itself to be applied to other campaigns, as well within the organization/realm of customer service providers. The only requirement, to validate it against other campaigns, would be the same input per campaign, knowing the dependent and independent variables and the description of the methods used to measure these variables. 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 as defined in section 2.1 were met. The data to information and knowledge traversal has been completed and resulted in a performance matrix which can be used to underline, plus support the decision making processes within an organization.

7.4. Reflections and limitations

During the writing of this research, a lot of (quality related) data still had to be entered into the systems, this resulted in long waiting times for the final resulting dataset, due to the limited availability of (quality related) datasets. This research's case-study has been based on one campaign/provider, qualitatively, this has been sufficient; however, quantitatively there are a lot more possibilities. For example, the historic results found could be extrapolated towards the current results and compared to other campaigns or service providers; however, this is very time-consuming and is out of scope for this research. The quality of the dataset could be further enhanced by eliminating external factors, such as incentives plus the impact of new personnel on quality and performance variables. This is a high frequent and ongoing process though and by doing so the risk on removal of useful data is high.

Another challenge in this research is the assumptions on variables that project managers have acquired even if (part of the) data is telling them otherwise or data that validates their assumptions is not available. Project managers do not have any quality related targets; they just have the responsibility for the service-level and conversion. Influences on these variables are often not perceived or ignored. The presentation of these influences can still be optimized so that managers understand the total cycle of quality and performance.

The goal of this research is finding key performance indicators from large quantities of data, while keeping the limitations of the research in mind, these indicators were found. The main step towards adaptation is the appliance of the rules found and implementing these into the systems, by the generation of decision trees and business rules on three different levels (formal, informal and technical). The results are easily applied within management overviews, as well as any technical systems that aid the decision making processes within the organization through using of existing terminology. The performance matrix, as well as the business rules as a result from the analysis, can be used or adapted as performance thresholds for the variables available in current systems and or management overviews. For example, in the event of an average or individual score for an independent variable dropping below the threshold derived from the business rule, the value could be flagged within the system or management overview. As a result of this, an exception could be sent using current available systems; furthermore, an overview of the data in a dashboard is required by the respondents. An example of such a dashboard using fictional data has been provided in Appendix E: Example Dashboard.

8. CONCLUSIONS

The following sections will provide the answers to the sub research questions by discussing the deliverables and by doing so pave the way to answer the main research question of this research, the research questions are repeated before the final concluding answer is provided in the accompanying text.

8.1. SUB-RESEARCH QUESTIONS

8.1.1. BUSINESS INTELLIGENCE

i. How can (business) analytics assist with the extraction, transformation and loading of data arranged within the area of business intelligence?

By the use of statistics (applied within the domain of data-mining), large amounts of data can be analyzed to find meaningful patterns and show information that would not have been accessible by analysis through other means, the application of business intelligence algorithms to the historical data used in this research allows data to be prepared for further analysis. In order to do so, access to the data needs to be gained. Furthermore, for application of business intelligence, a deeper understanding of the available data needs to be gained. One could say that without business analytics, the data and information to knowledge traversal is not possible. Business analytics allows for a deeper understanding of the data and optimization of the data structures. The extraction, transforming and loading of the data in turn is a logical follow up of the steps taken within the area of analytics.

ii. How can we define patterns from data using business analytics?

EXTRACT

By the use of data mining patterns which can be seen as the next step in the area of business analytics, or in other terms, the data to knowledge/wisdom traversal. As concluded in the first research question in this section, business intelligence is mainly used as a support measure in preparing the data for analysis. This means going through the available data and selecting it on its suitability for analysis. This implies the sifting of variables and following up on the relations discovered while analyzing the data in the business intelligence or analytics phase.

TRANSFORM

Transforming the data is performed by data cleansing and data preparation. It is one of the key steps in order to mine information from historical data. Essentially, deleting NULL values and values that do not hold up to the variables possible thresholds and the business rules applicable to the dataset, which are defined in the extraction and analysis phase. The transformation phase also includes the variable assessment and type transformation based on the requirements of the data mining technique, which has been selected for the analysis of the raw data. This is an ongoing process and takes place, even when during analysis certain data mining techniques require specific data types or input.

LOAD

Business intelligence in this research is meant to retrieve large amounts of data, showing and enabling analysis of historical results and aiming towards prediction of variables, as well as outcomes for dependent variables. The loading of data, the extraction of data, as well as the transformation of data is often performed by software applicable to the realm of data mining; however, it is by no means exclusively applicable to the data mining realm. The software used in this research (Rapid Miner) is harboring a combination of business intelligence and data mining functions.

ANALYZE

Using data mining technologies and algorithms, in order to obtain patterns from uncategorized data allows obtaining information and transforming this information into knowledge. Thus, this allows the application of knowledge in the process of decision making as shown in Figure 5 and Figure 6.

8.1.2. DATA MINING

i. Which techniques are applicable for mining historical data?

Most data mining methods suffice in order to retrieve information from historical data. The focus of this research lies within the pattern and rule based algorithms within the data mining realm. By the use of classification and clustering of variables, in combination with various techniques that are available in the area of extraction, transformation and loading of data, patterns arise from seemingly unconnected data. This research set out specific goals for the mined data in terms of prediction and description of the models that have been generated. The goals are of great influence when selecting the algorithms to be executed. This research made use of basic techniques provided with Rapid Miner, focusing on descriptive techniques for categorization (binning) and the predictive techniques for the definition of decision trees and business rules.

ii. How can we select performance indicators from historical data using data mining?

By the definition of most influencing variables it can be determined which indicators (variables) are prone to change the results of dependent variables. Literature study resulted in a long list of dependent variables; a short list emerged by comparing the long list to variables actually measured by the Customer Service Centre, these values have been calculated according to the specifications defined in the long list and were analyzed according to their influence on several dependent variables.

8.1.3. Key Performance Indicators

i. What are the required elements for a (key) performance indicator?

By literature study, it was found that any variable can behave as a key performance indicator based on the dependent variables measured. Using literature from the field several performance indicators and key performance indicators are found and these in turn were used to determine the influences these variables force on another. When looking at the previous statement, it is found that the main element for a performance indicator is the level of influence on another variable. These influences can be measured and the strength of the influences is of big importance when defining the key performance indicators within an organization or process.

ii. What makes an indicator a key performance indicator?

The definition used by this research for a key performance indicator is the historical and future tracking of measurements, which enable an organization to perform or excel at their most fundamental levels. Sections 6.2.2 and 6.6.1 show the key performance indicators found during this research and the case study performed. As stated by the answer on the previous research question, the most important relevant factor is influence, how does an (performance) indicator influence financial or operational measured variables. By the definition and investigation of these influences key performance indicators may be found.

iii. Which (key) performance indicators can be derived from the historical data available?

This research found dependent variables by literature study and cross compared these with the variables already measured on an operational level thus resulting in the key performance matrix detailing the indicators that are crucial in obtaining performance goals, the final key performance matrix is displayed in table 11.

8.1.4. Business Rules

i. What are the required elements for a business rule?

The building blocks for a business rule describe input, output and an effect, for an effect to take place there needs to be input in the form of an event, as well as output (essentially the affected event). While this is comparable to the description of a process there are differences, for example, thresholds and the effects of an event on another variable or process are needed to construct a business rule.

ii. How should business rules be presented to an analyst or system?

When presenting business rules to a system, it is crucial to uphold a closely defined syntax (minute naming of variables), when presenting business rules to an analyst, natural language should be used. Next to lingual properties, business rules can also become too detailed for an analyst to understand its meaning or not have enough detail for a system to draw solid decisions from the rule. The key understanding in determining the presentation of rules is to determine the use and application. When determining the presentation, it should be kept in mind that business rules and their effects can be broken down in to pieces. In addition it could theoretically be possible to construct one business rule that summarizes all effects, exceptions and thresholds for the process or organization. The key in determining the presentation of the rule is keeping it simple and understandable for the end-user and the person that is implementing these rules into the systems.

iii. Which business rules can be derived and aligned to the previously mined data?

When looking at the case study performed, both informal and technical business rules were found, in order to traverse these to a formal level, calculations of (key) performance indicators should be included. It should be noted that business rules with a technical description can also be implemented in software that support decision making processes.

8.2. Main 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 within a customer service center?"

By combining the 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, as well as the weight of impact they have as shown in figure 35. 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 Rule Extraction Matrix 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 Rule Extraction Matrix 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.

Figure 35: Employee performance versus organizational performance

8.3. Further research

This section describes possibilities for further research.

The research presented to the reader allows several possibilities for further research. Firstly, the method can be applied to other business areas/cases. Next to the application of the method, the method can also be extended with steps that concentrate on the verification of the results at hand. Currently, this is performed by taking expert interviews as a means of evaluation; however, it would be suitable to include multiple campaigns and extend the model to other business areas.

As discussed in the previous section, the application of historic results (prediction) can also be compared towards future results, thus verifying the generated models. Finally, the research results can be used to further extend the key performance indicators used in standards like COPC.

9. References

Aertsen, J. (2010, 02 18). *Extract transform load*. Retrieved 05 12, 2010, from Business intelligence compendium: http://www.businessintelligencecompendium.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.

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

Cameron, K. S., & Whetten, D. A. (1983). Organizational effectiveness: A comparison of multiple models. In K. S. Cameron, & D. A. Whetten, *Organizational effectiveness: One model or several?* (pp. 1-24). New York: Academic Press.

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.* Retrieved 05 20, 2010, from CRISP-DM Consortium: http://www.crisp-dm.org/CRISPWP-0800.pdf

Cheong, K., Kim, J., & So, S. (2008). A Study of Strategic Call Center Management: Relationship between Key Performance Indicators and Customer Satisfaction. *European Journal of Social Sciences , 6* (2), 268-276.

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.

Corbin, J., & Strauss, A. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park: Sage.

Darrow, B. (2003, February 3). Making The Right Choice. *Computer Reseller News* , p. 16.

Deming, W. E. (1986). *Out of the Crisis.* Cambridge: MIT Center for Advanced Engineering Study.

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*. Retrieved 04 28, 2010, from 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.

Hart, H., & Boeije, H. (2006). *Onderzoeksmethoden.* Utrecht: Boom Onderwijs.

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

Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard - Measures that drive performance. *Harvard Business Review* , 70-79.

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

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

Lytras, M. D., Russ, M., Maier, R., & Naeve, A. (2008). *Knowledge Management Strategies: A Handbook of Applied Technolgies.* New York: Hershey.

Mazlack, L. J. (2003). Causality Recognition For Data Mining In An Inherently Ill Defined World. *BISC FLINT-CIBI International Joint Workshop On Soft Computing For Internet And Bioinformatics.* California: University of California.

Merriam, S. B. (1998). *Qualitative research and case study applications in education.* San Francisco: Josey-Bass.

Morgan, T. (2002). *Business rules and information systems: aligning IT with business goals.* Boston: Addison-Wesley.

Morris, H. (2003). The Financial Impact of Business Analytics: Build vs. Buy. *DM Review , 13* (1), 40-41.

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

Nelson, M. (2010, 04 30). *What are the key components of a key performance indicator*. Retrieved 06 16, 2010, from Ibis Associates: http://www.ibisassoc.co.uk/key-performance-indicators.htm

Perry, D. E., Sim, S. E., & Easterbrook, S. M. (2004). Case Studies for Software Engineers. *Proceedings of the 26th International Conference on Software Engineering (ICSE'04)*, (pp. 736-738).

Piatetsky-Shapiro, G. (1991). Knowledge Discovery in Real Databases: A report on the IJCAI-89 Workshop. *AI Magazine , 10* (5), 68-70.

Rockart, J. F. (1978). *A new approach to defining the chief executive's information needs.* Massachusetts Institute of Technology, Center for Information Systems Research. Cambridge: Alfred P. Sloan School of Management.

Rossman, G., & Rallis, S. (2003). *Learning in the field: An introduction to qualitative research.* London: Sage.

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science , 33* (2), 163-180.

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

Sayad, S. (2012). *An Introduction to Data Mining*. Retrieved 04 12, 2012, from Faculty of Applied Science and Engineering: http://chem-eng.utoronto.ca/~datamining/dmc/data_mining_map.htm

Seifert, J. (2004, 06 13). *Data Mining: An Overview.* Retrieved 6 13, 2010, from Congressional Research Service: www.fas.org/irp/crs/RL31798.pdf

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.

the Business Rules Group. (2000). *Defining Business Rules: What are they realy.* Guide.

Tijssen, R., Spruit, M., van de Ridder, M., & Raaij, B. (2010). BI-FIT: Aligning Business Intelligence end-users, tasks and technologies. In M. Cruz-Cunha, & J. Varajão, *Enterprise Information Systems Design, Implementation and Management: Organizational Applications* (pp. 162-177).

Verschuren, P., & Doorewaard, H. (2000). *Het ontwerpen van een onderzoek.* Utrecht: Lemma B.V.

Vleugel, A., Spruit, M., & van Daal, A. (2010). Historical data analyisis 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.

10. APPENDICES

10.1. APPENDIX A: IN-DEPTH SEARCH-PLAN

The in-depth search-plan shows which literature and subjects will be looked on, the research subjects have been defined as business intelligence, data mining, key performance indicators and business rules. Per subject an item is described with keywords that have been queried in order to obtain the proper knowledge for researching the various domains and answering research questions.

i. Business intelligence
   a. Knowledge cycle and knowledge management
   b. (Open source) business intelligence tools
   c. Business process management
   d. Decision support systems
   e. Scorecarding ((automated) balanced scorecard) and performance measurement

ii. Data mining
   a. Methods applicable to data mining
   b. (Open source) data mining tools
   c. Data/case selection
   d. Input/output mechanisms
   e. Live monitoring versus data storage
   f. Association rules

iii. Key performance indicators
   a. Structural definitions
   b. Smart measurements for key performance indicators
   c. Thresholds and monitoring

iv. Business rules
   a. Structural definitions
   b. Required elements
   c. Alignment information technology
   d. Definition of business rules (process steps)
   e. Rule discovery

10.2. Appendix B: Data structure

10.2.1. Data structure VANAD

This section details the available data structures that are stored with VANAD, in this case asterisk_cim and quality_monitoring. Figure 36 below displays the data-structure and relations between the data available for this research. The relation lies between asterisk_cim_asterisk_uid and quality_monitoring_unique_id.

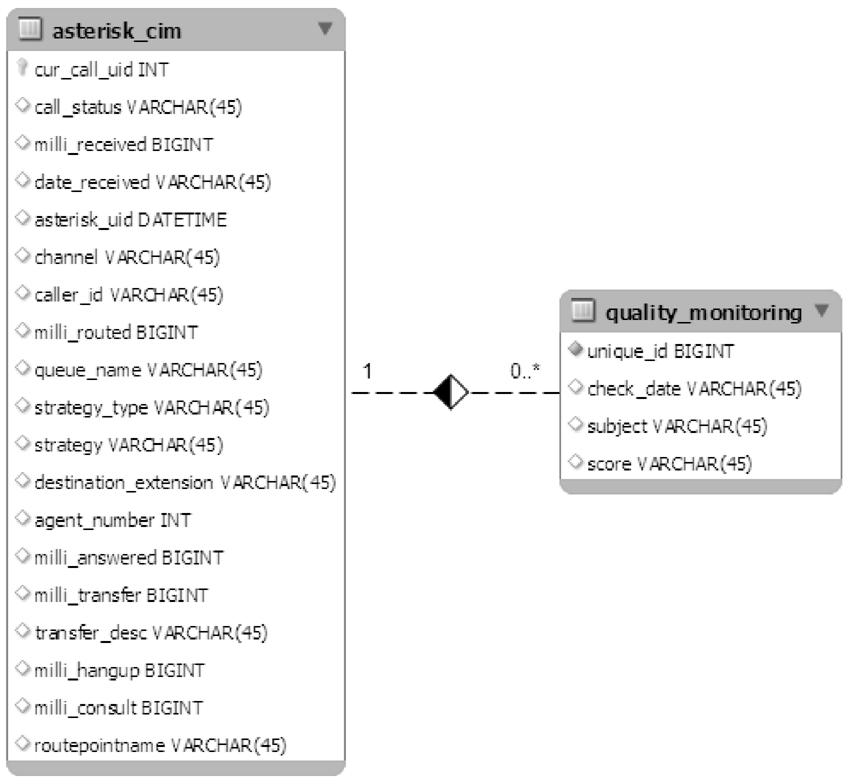

Figure 36: Data-structure of data sources VANAD

10.2.1.1. Asterisk_cim items

Asterisk_cim contains all call-aspects, it is extracted from the main database which handles the relationship between the quality monitoring database stored at VANAD and the customer database stored with the customer.

table_descriptors	metadata
cur_call_uid	Randomized ID number,
call_status	Status call
milli_received	The moment a call is received displayed in milliseconds following unix_timestamp (amount of milliseconds since 1-1-1970).
date_received	The moment a call is received displayed in YYYY-MM-DD HH:MM:SS.
asterisk_uid	The unique_id used by asterisk to identify a consumer's call, this ID is also used in the quality_monitoring database.
channel	The asterisk_channel a call was relayed towards.

caller_id	Consumer's telephone number.
milli_routed	The moment the decision is made to route a call towards a queue or an agent.
queue_name	The queue number.
strategy_type	Used routing strategy/type, this shows if a call has been in a waiting queue or if the call's been transferred to an agent.
strategy	Routing strategy descriptor.
destination_extension	The number of an agent's extension.
agent_number	The agent's number.
milli_answered	The moment a call is received by an agent displayed in milliseconds following unix_timestamp (amount of milliseconds since 1-1-1970).
milli_transfer	The moment a call is forwarded by an agent to an external number displayed in milliseconds following unix_timestamp (amount of milliseconds since 1-1-1970).
transfer_desc	The number a call is forwarded towards
milli_hangup	The moment a call is hungup either by the customer/agent/system displayed in milliseconds following unix_timestamp (amount of milliseconds since 1-1-1970).
milli_consult	The moment an enquiry call takes place displayed in milliseconds following unix_timestamp (amount of milliseconds since 1-1-1970).
routepointname	The queue's name.

TABLE 13: ASTERISK_CIM META DATA

10.2.1.2. QUALITY MONITORING ITEMS

Quality monitoring is an export of the performance variables the quality monitoring team is able to check a call, these performance variables exist of 13 properties such as listening capabilities and how was the solution towards a customer presented.

table_descriptors	metadata
unique_id	ID number,
check_date	Timestamp checked service requests
subject	One of 13 unique call descriptors
score	The sore for the subject on a scale from 1 to 5

TABLE 14: QUALITY_MONITORING META DATA

The following takes place, asterisks stores service requests ID's which are identified through a unique ID in essention the unix timestamp followed by a random number. When service requests are monitored the unique ID (unixtimestamp.randomnumber) is used as an identifier for the monitored service requests. In the current situation not every service request is monitored and thus the relationship is 0..*. The software responsible for relaying call towards the quality monitoring team aims to provide the quality monitoring team with at least one service request per month per employee.

The thirteen quality monitoring call descriptors are listed below:

- Administratieve afhandeling

Administratieve behandeling is a review on the translation of a service requests to a registration of the service request in the systems of the service provider. The subject score is an answer to the following question: “In what level are the customer’s needs/wishes translated towards the actions available within the systems of the service provider?” (consumer upgrades/requests for support etc.).

- Afsluiting

Afsluiting treats the subject of how the customer is treated in terms of ending the service request. The subject score provides an answer to the following question: “In what level does the end of the service request abide to the standard set out by the service provider for ending a service request?”.

- Behoefte(n)bepaling

Behoeftebepaling reviews the identification of the needs of a customer. The subject score is an answer to the question: “In what level did the employee determine the needs of the customer and did the employee ask the right questions following the guidelines handed out by the service provider?”.

- Gespreksvoering

Gespreksvoering is a review on how the customer is treated in terms of service request and natural language. The subject score provides an answer to the following question: “In what level are the rules provided by the service provider on service request language followed?”.

- Introductie

Introduction treats the subject of how a service request is introduced to the customer. The subject score is an answer to the following question: “In what level is the service request introduced in conformity with the rules provided by the service provider?” (subject introduction, naming the customer and organization).

- Klanttevredenheid

Klanttevredenheid reviews the identification of the satisfaction of the customer. The subject score is an answer to the following question: “In what level did the employee poll the customer’s satisfaction?”.

- Leiding behouden

Leiding behouden is a review on how the service request is guided. The subject score provides an answer to the following question: "In what level did the employee control/guide the service request?".

- Luisteren

Luisteren treats the subject of how the employee heeds the details in a service request. The subject score is an answer to the following question: "In what level is an employee able to translate minute specifics into actionable information?".

- Productkennis

Productkennis reviews the employee's knowledge towards the product of the service provider. The subject score provides an answer to the following question: "In what level is an employee able to answer the questions of a consumer without external help and is the information provided by the employee accurate".

- Propositie(s) en Analyse

Propositie(s) en analyse is a review of the employee's offer to the customer. The subject score is an answer to the following question: "In what level is the offer of a solution by the employee in accordance with the needs of the customer".

- Samenvatting

Samenvatting treats the subject of how the employee summarizes the service request and it's details. The subject score provides an answer to the following question: "In what level does the employee summarize the customer's needs and possible solutions to those needs?".

- Structuur

Structuur reviews the subject of how the service request is structured. The subject score is an answer to the following question: "In what level is the service request structured in terms of the following hierarchy: "introduction, analysis, proposition, closure"?".

- Vragen

Vragen is a review of how the employee allows the customer to ask questions. The subject score provides an answer to the following question: "In what level does the employee provide the customer with the opportunity to ask questions?".

10.2.2. DATA STRUCTURE SERVICE PROVIDER

This section details the data structures stored with the service provider in this case the performance_data. Figure 37 below displays the data-structure of data sources available with the service provider. The figure shows the output that is monitored weekly and compared between the various customer service centers.

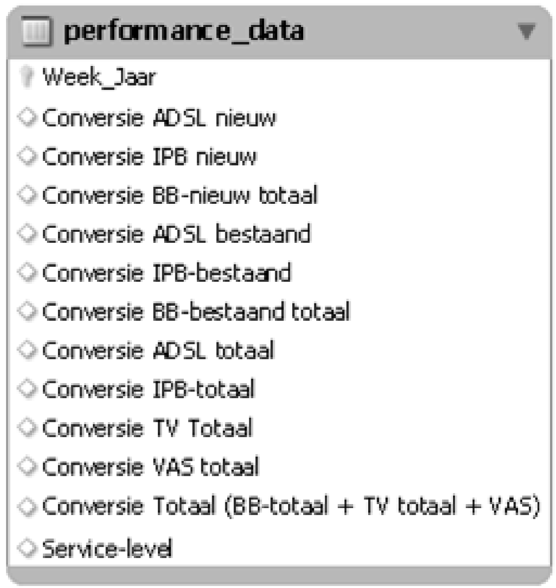

FIGURE 37: DATA-STRUCTURE OF DATA SOURCES SERVICE PROVIDER

table_descriptors	metadata
week_Jaar	Date
conversie %	Percentage of accepted propositions (Totaal is important)
service-level	Calls answered within 20 seconds + Calls abandoned within 20 seconds)/(Total calls answered + Total calls abandoned)

TABLE 15: PERFORMANCE_DATA META DATA

The above table details the structure used by the service provider for weekly analysis between custumor service centers and their overall performance. Through the usage of the date details these data can be used to evaluate the performance levels measured internally at the customer service center towards the external measurements performed by the service provider.

10.3. Appendix C: interview questions

i. What would you deem to be the biggest obstacle in carefully monitoring performance within your campaign?
ii. How would you consider a project being well managed?
iii. What is your gathering on the predictions and descriptions of the results presented?
iv. How would you further optimize performance with the results presented?
v. Looking at historical results, how do you feel the accuracy of the results can be improved?
vi. How would you like the results to be displayed/deployed?
vii. Which analysis has taken place in the past concerning the key performance indicators that are detailed in the results?
viii. How would you base your decisions on the information presented to you?

10.4. Appendix D: interview response

10.4.1. Respondent One (Manager)

1. Respondent one mentioned the communication with the underlying levels of management as one of the key factors that influence performance monitoring, next to this there is no fixed key performance indicator that defines quality neither a requirement in business reviews performed by the customer and project manager. Ideas concerning quality are outlined in the strategic decision making process; however, on a tactical level these are often neglected since the main focus lies on operational performance indicators in contrast to quality indicators.
2. Respondent one feels a project is well managed when the perception and expectations of a customer service provider are met, the agreements on the topic of quality and performance should be met, as well as response times concerning questions of the customer service provider should be low. Two critical topics lie at the basis of this relationship, the transparent and timely notification towards the customer that targets will not be met and the adjustment of internal processes in order to provide better conclusions towards a customer.
3. Respondent one says the results provided by this research show the relations that were not available in the past, praising the availability of these relations to enhance his project's performance. Next to the quality and performance relation, the respondent looks at the motivational aspects that these results can have on an employee, in order to further increase the performance, employees should be offered a training schedule based on their strengths and weaknesses.
4. Respondent one defines the need for commitment of a customer towards increasing quality, without customer understanding of the relationship quality and performance no budget will be made available to further increase the quality indicators. Respondent one also mentions the need for translation of these relations towards the tacit and operational levels, the level of acceptation and realization is currently low, targets towards the tacit and operational levels can help drive quality and performance.
5. Respondent one feels the accuracy can be affected by constantly monitoring the relationship between quality and performance and verify this by comparing the results over multiple campaigns, verification should occur based on the comparison of historical and current records.
6. Respondent one would like to see a graphical overview of the relations which in turn is based on the decision trees presented to the respondent. The customer should receive numerical overview of the relations based on the graphical overview.
7. Respondent one notes that currently performance and quality monitoring have not been used in this level of detail, in the past a

random sample selection has been used to verify performance which is further supported by batch analysis of work-stock.

8. Respondent one defines the aim for driving performance and quality should be commitment with the customer based on combined decision making driven by transparent information concerning the relations between quality and performance. The results presented can be used to further drive performance by presenting these towards the customer service provider and together influence the current decision making processes, as well as internal processes.

10.4.2. Respondent Two (Analyst/manager)

1. Respondent two feels the biggest obstacle is the application of theoretical solutions in practice, this is often influenced by sluggish communication and priorities set by lower levels of management. Another obstacle is managing the customers perception and expectations, how to present relations between costs, performance and quality.
2. Respondent two defines the need for theoretical grounding of key performance indicators in terms of positives and negatives, key performance indicators that have not met their performance thresholds should be evaluated and used to alter the customers' internal solutions and propositions.
3. Respondent two comments on the relations, determining that these are confirming assumptions on the relation between quality and performance, showing these decision trees to a customer can help in finding budgets to invest in performance.
4. Respondent two identifies the need to continually calibrate key performance indicators with the service provider and optimize internal processes in order to make employees aware of the impact of their actions on the overall performance.
5. Respondent two defines the need for using more data sources such as external monitoring and calibrating these based on outlier mechanisms. The respondent identifies the need for the application of the data mining solutions towards multiple campaigns, conclusions are now drawn on one campaign, for them to be applied companywide and multiple campaigns should undergo a thorough analysis.
6. Respondent two feels in order to optimize performance using the results the key performance indicators should be presented to the end users in real time, using monthly/two monthly evaluations with the customer to determine strategies that further optimize performance.
7. Respondent two comments on the past performance monitoring as follows, the internal and external performance was measured using a two-stage model, monitoring performance was performed internally, as well as externally by taking random samples of communication and

comparing these to the analysis performed by the customer service provider.

8. Respondent two identifies two uses for the results, first the results help/support the qualitative analysis of decisions and allows secondly to determine which areas should be targeted first in terms of quality and performance.

10.5. APPENDIX E: EXAMPLE DASHBOARD

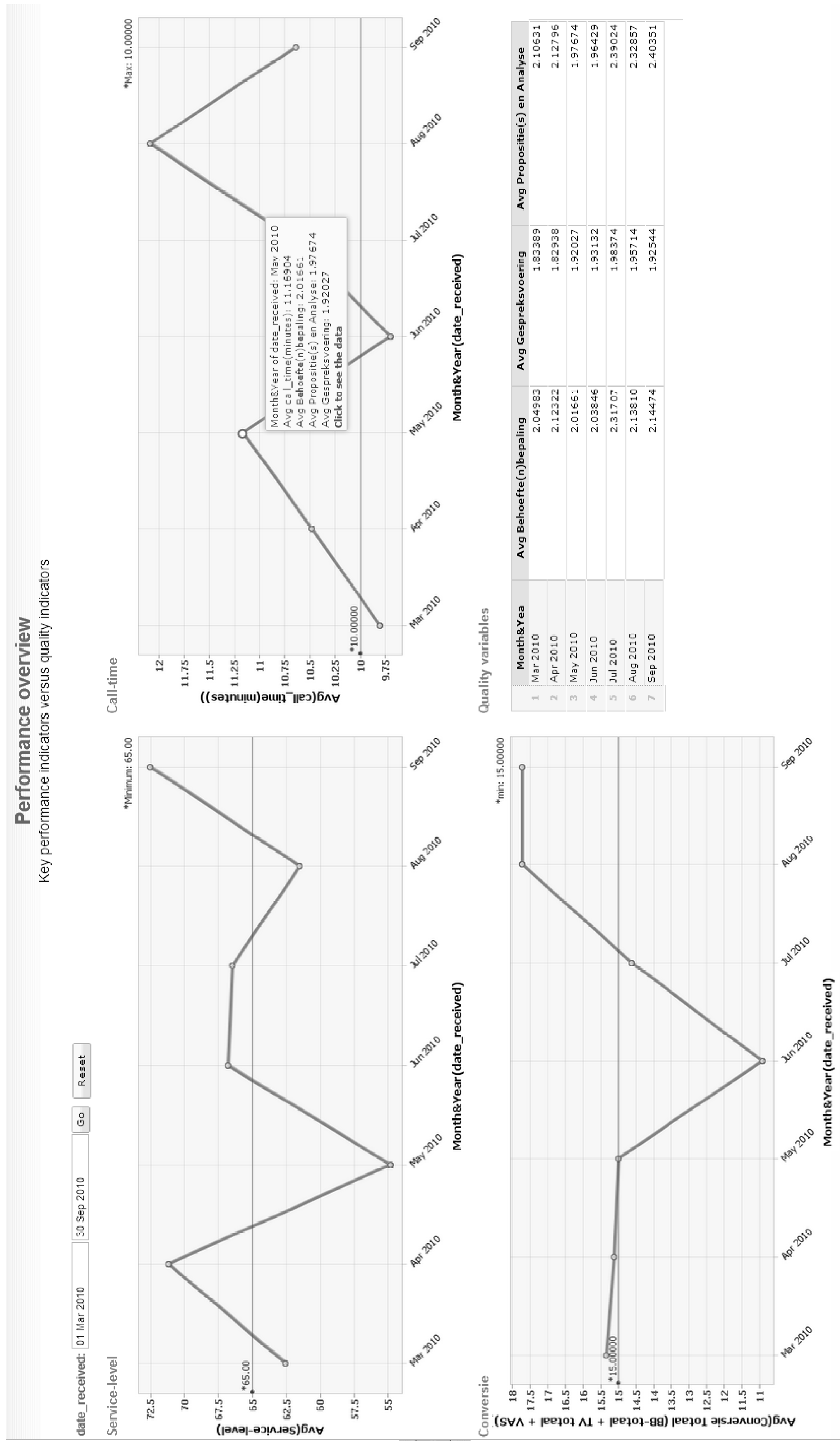

FIGURE 38: EXAMPLE PERFORMANCE DASHBOARD