# Determing the Business Value of Business Intelligence with Data Mining Methods

Karin Hartl, Olaf Jacob

Department of Information Management
University of Applied Sciences Neu-Ulm (HNU)
Neu-Ulm, Germany
karin.hartl@hs-neu-ulm.de, olaf.jacob@hs-neu-ulm.de

*Abstract*—**This paper presents a research project which aims to determine the value of Business Intelligence (BI) and Corporate Performance Management (CPM) with the help of Data Mining methods. The starting point of the research is the hypothesis that the value proposition of BI can be measured on the success of CPM. Previous empirical studies try to define the impact of BI on CPM with research methods like explanatory factor analysis or structural equation modeling. This paper discusses the use and benefit of Data Mining methods for exploring the value of BI. It clarifies why specific Data Mining methods are seen as a beneficial tool to determine the relation between BI and CPM.**

*Keywords-Business Intelligence; Corporate Performance Management; Data Mining; Business Value.*

## I.    INTRODUCTION

The challenge companies have to face nowadays for success and existence proves to be increasingly difficult. Globalization intensifies the competition and digitalization leaves enterprises with an immense amount of mainly unstructured data. These data and the contained information, however, are assumed to be the key to ensure the survival of an enterprise in the rapidly changing business environment. BI as a method of analyzing data and the business environment promises companies to support their decision making process [2]. The support is achieved by acquiring, analyzing and disseminating information from data significant to the business activities [1]. Accordingly, BI is seen as a source for quality data and actionable information. This implies that the appropriate use of BI systems supports the success of organizations [3].

As BI projects are not exempt from the increasing pressure in companies to justify the return on IT Investment, the business value of BI needs to be measured [5]. Due to the abstract nature of BI capturing its value, it is a strategic challenge [6][4]. Generally, BI systems don't pay for themselves strictly by cost reduction. Most BI benefits are intangible and hard to measure [5]. Williams and Williams [6] point out that the business value of BI lies in its use within the management processes. Therefore, the concept of CPM evolved, which is understood as the appropriate context to prove the value proposition of BI [4]. It is defined by Gartner as "*an umbrella term that describes all processes, methodologies, metrics and systems needed to*

*measure and manage the performance of an organization"* [9]. CPM presents the strategic deployment of BI solutions and is born out of a company need to proactively manage business performance [8][10]. Inferentially, CPM needs BI to work effectively on accurate, timely and high quality data and BI needs CPM for a purposeful commitment [9]. Consequently, it is expected that the effectiveness of CPM increases with the effectiveness of the BI solution and therefore company success improves as well [11]. A hypothesis can be put in place, which states that the value proposition of BI can be measured on the success of CPM. This research aims to define the link between CPM and BI with the use of Data Mining methods and is based on the findings of Jacob and Lien Mbep [3]. In the research field of BI, exploratory factor analysis and structural equation modeling are the dominant research methods. As addressed in this paper, it is assumed that Data Mining techniques are able to answer different kinds of research questions than the above mentioned approaches on the subjects of BI and CPM. Data Mining could be suitable to gain more detailed information and could be a more appropriate approach to examine the business value of BI on the effectiveness of the CPM processes.

In Section 2 of this paper an overview of subject related research and its importance for this research is given. Section 3 highlights the motivators for using Data Mining techniques to discover the relationship between BI and CPM. In Section 4, the aspired research approach is explained including the Data Mining methods which will be applied. Section 5 closes with a short conclusion on why Data Mining is seen as a beneficial approach to determine the business value of BI.

## II.    SUBJECT RELATED RESEARCH

In the last couple of years, various studies regarding the business value of BI emerged. An early approach has been made by the Viva Business Intelligence Company [4] in discussing general principles on how the business value of BI could be measured. The article underlines that the direct monetary benefits are hard to calculate and that the significance of BI programs lies in the production and analyzation of information [4]. Even though the standardization of the BI-output for measurement purposes is suggested and possible measures are stated, no detailed

examination has been accomplished. Williams and Williams [6] expose the necessity to determine how BI is used in a company for the quest of defining the value proposition of BI. It is shown that the business value of BI lies especially in its use within the management processes of a company that impact the operational processes. The return on BI investment is assumed to be measurable on the increased revenues and reduced costs. But BI is more than monetary benefits. It is a complex process that makes an in depth evaluation of the impact it has on the management processes necessary. In 2004, Miranda [12] brought BI into context with CPM by summarizing CPM as a business management approach that supports companies in the way they operate by using business analysis. CPM is a management process based on BI systems. Therefore, CPM is identified as a suitable framework for determining the business value of BI. This conclusion provides the foundation for more detailed research in the field, including the following.

Empirical studies on the investigation of the business value of BI have mainly been realized just recently. Yogev et al. [13] addresses the question of the business value gained by implementing a BI system in an enterprise through using a process oriented approach. The research model formulated is built on the resource based view of a firm. It identifies key BI resources and capabilities as possible explanatory factors of the value creation that can be accomplished with the implementation and application of a BI system. Hypothesis are formulated that can be summarized to state that BI has a positive effect on the operational and strategic business processes. Data have been collected using a survey consisting of seven-point Likert scale items, anchored at the ends by "strongly agree" and "strongly disagree". The research method applied is structural equation modelling and the confirmatory factor analysis is identified as showing a satisfactory model fit. The results illustrate that BI has a positive effect on both the operational and the strategic level of the company. Nevertheless, no further details are given on the intensity of the positive effect BI has on performance management, on how this positive effect can be measured nor on the BI related resources which create these positive effects. Richard et al. [11] are the first to investigate the impact of BI on CPM. The aim of the study is the examination of the impact of commonly used BI technologies on the CPM related management practices which include planning, measurement and analysis. The role BI plays in supporting CPM related managements practices is to be identified to enable IT practitioners to better understand the influence of BI technologies across the CPM cycle. The main research hypothesis states that *"the more effective the BI implementation, the more effective the CPM-related management practices"* [11]. The research model as shown in Figure 1 supposes that BI directly influences and supports measurement, planning and analytics. The effectiveness of planning, measurement and analytics, again, influences the effectiveness of the company processes. To answer the

research hypothesis, sample data have been collected by using a questionnaire which is based on items in the Likert scale format [11].
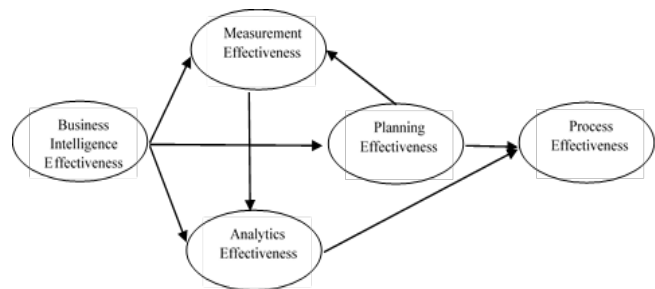


Figure 1.    Research Framework [11]

After collecting the data, Richard et al. [11] used an exploratory factor analysis to reduce the number of variables complied in the questionnaire followed by the partial least square (PLS) analysis. The findings suggest that BI positively influences planning effectiveness, analytics effectiveness and through these, indirectly influences process effectiveness as well. Even though the research identifies a direction of how BI influences CPM, the specific BI mechanisms who do so are not defined. Therefore, Richard et al. [11] suggests further studies on this subject, which will be done by this research.

The subject related work is complemented by this study as the previous findings have been used as the initial point. Miranda [12] identifies the importance CPM has in determining the business value of BI and Richard et al. [11] applies this knowledge. But besides discovering and proving a positive connection between BI and CPM, both researches lack detail. It is still not clarified which BI related tools and processes influence the CPM of a company and which CPM processes are effected. Therefore, the CPM process has to be identified in more detail. Based on a process model defined by Lien Mbep et al. [10] and an empirical analysis Jacob and Lien Mbep [3] identified a set of both CPM and BI related items. These items are understood as suitable to measure the appropriate use of BI on one site and CPM on the other. By bringing them together it is aimed to identify the BI related resources and factors which influence the CPM cycle positively. Furthermore, the strength on how the resources and factors positively influence the effectiveness of a company CPM is to be identified.

## III.    MOTIVES FOR THE DATA MINING APPROACH

The common approach in the research field of determining the value of BI is using exploratory factor analysis first and then confirmatory factor analysis second. With the exploratory factor analysis correlating items are organized together in groups and summed up as a factor. Data can be structured and reduced this way. This structured and reduced data are then analyzed with the PLS method by seeking the optimal predictive linear relationship to assess the previous defined causal relationship. The creation of

factors for compacting information might be the right approach for many research subjects, but is it the only correct approach for defining the business value of BI? It is assumed that Data Mining can highly contribute to the subject. It presents the opportunity to answer different as well as more specific research questions then the commonly used approaches. Instead only testing assumed hypothesis with Data Mining otherwise undiscovered data attributes, trends and patterns can be explored [14]. This can be done with predictive and explanatory Data Mining methods. With explanatory Data Mining data can be interpreted and a better understanding of connections in the dataset is to be achieved [18]. Predictive Data Mining procedures aim to define prediction models from existing data, which allows forecasting unknown values in new data [18]. Although Data Mining is only seen as most suitable for large datasets, Natek and Zwilling [17] disclose that the use of small datasets in specific Data Mining analysis are not limiting the use of the tool. Data Mining can be understood as an extension of statistical data analysis and statistical approaches [16]. Both approaches aim to discover structure in data, but Data Mining methods are generally robust to non-linear data, complex relationships and non-normal distributions [15]. These differences between Data Mining and the commonly used statistical approaches are assumed to supply more detailed and surprising results for the research field of BI and CPM.

## IV. RESEARCH APPROACH

This research bases on the findings of Jacob and Lien Mbep [3], where a set of criteria that is seen as most suitable to represent CPM on one and BI on the other side has been identified. In total a set of 20 CPM related items and 28 BI related items have been selected. Now a supplementary study is needed to bring the criteria of both fields together and to explain the relationship between BI and CPM [3]. Therefore, the identified criteria have been transformed into questionnaire items which are to be answered on a five-point Likert scale. The anchor points at the ends of the scale are "does not apply" and "does apply" and an additional definition "applies half and half" for the mid stage has been defined. The data collection is taking place at the moment by using an online questionnaire. Subjects are German companies who use BI for supporting their performance management. Hence decision makers from management, controlling and IT are addressed. With the survey results inter alia the following research questions are to be answered which can be done by using various Data Mining techniques.

*R1: Which relations persist between the CPM criteria and the BI criteria?*
*R2: Which patterns exist between the interviewed companies?*

*R3: Is the CPM development of a company predictable through the occurrence of the BI characteristics of a company?*

To answer the above questions a procedure model has to be first defined. In the Data Mining literature, a broad variety of those can be found, like the Cross Industry Standard Process for Data Mining (CRISP-DM) process model or the overall procedure model Knowledge Discovery in Data Bases (KDD) [18]. They all have the main steps in common. The ones shown in Figure 2 will be followed in the examination of the defined research questions [18]. As the data are especially generated for the research purpose, no selection of the appropriate dataset is necessary. Therefore, the starting point for the data analysis will be preprocessing of the data. Data will be cleansed and missing, as well as conflicting values corrected. The main issue this research assumedly has to deal with are missing values. Cleve and Lämmel [18] suggest alternatives for dealing with missing values depending on the data structure. The important items of the questionnaire are formatted as Likert scale items and can be interpreted as metric data. Metric data can be preprocessed by replacing the missing values in the sample by the mean value of all item-based compiled answers. The mean values also can be stated by contemplating the datasets closest to the dataset with the missing value. This idea follows the k-nearest neighbours (kNN) approach and will most likely be applied to the collected dataset. After preprocessing, the data will be transformed in the required format for the applicable Data Mining technique. Data Mining algorithms demand specific data types. For example, the kNN prefers metric data and the Apriory Algorithm needs binary data [18]. Before operating, the below discussed Data Mining techniques it has to be ensured, that the correct data types for each method are provided. In the third step the data will be mined by applying the algorithm identified as most suitable to answer the above mentioned research questions. Afterwards, in step 4, the outcomes will be interpreted and evaluated.



Figure 2.   Research Procedure Model

As the study aims to determine the business value of BI on the processes of the CPM of a company, firstly, the relations between CPM and BI are to be explored. The common approach would be to correlate the collected data. This way, the linear connections between BI and CPM can be explored. To get a detailed result on criteria basis that would mean to correlate the 20 CPM items of the questionnaire with the 28 BI items, if no hypothesis are put in place beforehand. Accordingly, researchers would apply an exploratory factor analysis to reduce the data. This may mean more structure and a better overview, but also is associated with data loss and loss of accuracy plus detail.

Data Mining as a technique to discover new and unexpected patterns and relationships in data is assumed to be a second approach for determining connections and associations. In comparison with correlation or regression analysis many Data Mining techniques do not imply connections in advance but discover them automatically. It is assumed that the above mentioned research questions can be answered by the following Data Mining techniques, as shown in Table 1.

To answer research question one R1, Association Rule Discovery will be applied. With association rules, co-occurrence relationships between data items can be discovered, taking into account as many research items as needed and available [18]. This, indeed, can lead on the upside to more detailed results and on the downside to an enormous amount of discovered association rules. Unmanageable amounts of association rules easily can be organized by instating metrics that measure the interestingness of the discovered connections. An appropriate metric could be *Lift* [18]. To generate association rules many algorithms are available. The Apriory Algorithm is the classic procedure and works in two steps [19]. First, frequent itemsets are identified before the confident association rules are generated. In this research, association rule discovery will be applied to find relation rules between BI and CPM. Before applying the Apriory Algorithm to the compiled dataset the data will be transformed into binary variables. This transformation gives each of the items the two characteristics "distinct" or "not distinct". After executing the Apriory Algorithm to the cleansed data it is assumed that many anticipated connections between BI and CPM are shown. But also interesting new results are expected. As the analysis will include all criteria from both CPM and BI, detailed outcomes are targeted. The results will be association rules, showing which BI items and which CPM items appear together in one of the above mentioned characteristics. For example, a discovered rule could state, that if BI item 1 and BI item 2 are distinct, then there will be a high chance that CPM item 3 as well is distinct. Furthermore, it can be reasoned that the investment of a company in the development of BI item 1 and 2 results most likely in a higher CPM stage of maturity.

The second research question R2 asks for similarities between the companies interviewed. The regarding results could provide information that states if the company size influences the successful use of BI and CPM. Also it could be shown whether companies with well-established CPM strategies also have a well-functioning BI system. Corresponding results may create room for conclusions of a positive connection between the successful use and implementation of the BI system in a company and an effective CPM. Patterns and groups in the research criteria can be found by using clustering. Clustering organizes data without previous knowledge of potential groups [18]. It organizes the examination objects by means of their similarity.

TABLE I.      OVERVIEW OF THE APPLICAPLE DATA MINING TECHNIQUES

| Research Issues and applicable Data Mining Techniques | | |
|---|---|---|
| *Research Issue* | *Data Mining Method* | *Algorithm* |
| Examining the relationship between the CPM criteria and the BI criteria | Association Rule Discovery | Apriory Algorithm |
| Pattern discovery | Clustering | k-means algorithm |
| Predictability of the CPM development | Classification | decision tree modelling |

The objects belonging to one group are as much as possible homogenous based on their characteristics [18]. The groups, however, are as heterogeneous as possible among themselves [18]. In clustering all attributes available can be used in parallel. This offers a detailed view of the cluster features and enables a thorough view on the relations between BI and CPM. Clustering is done by defining a similarity and distance measure, which is also known as proximity measure. For ordinal scaled variables the *City-Block-Metric* is an appropriate measure [20]. A first interesting result on the research data is believed to be accomplished by using the k-means algorithm as it is the best known partitioning algorithm [21]. The data are iteratively partitioned into $k$ clusters by using a distance function. The quantity of clusters $k$ has to be defined beforehand [21]. A hierarchical cluster analysis, like the Ward's method, can help to define the number of clusters needed [21]. Results could be two or more clusters in which the companies surveyed can be divided. An example outcome could show that companies belonging to the same industrial sector group together in one cluster. The development of the BI items and CPM items in this cluster, furthermore, help to explore the usage and connection of BI and CPM in Germanys companies today. This facilitates the understanding of the connection BI and CPM have in Germany.

To evaluate the predictability of the CPM development in a company on the occurrence of the BI characteristics, as asked in question R3, decision tree modelling [18] can be applied. The decision tree learning is known as very effective and therefore a widely used technique for classification [21]. It is a hierarchical classification model which means that the research items are tested separately according to their importance. In that way, the possible classes are limited gradually and visually presented as a decision tree [19]. This is usually done by identifying successively homogeneous groups in a training dataset in concerning the classification variables [19]. The data collected will be used as training dataset. The decision tree results support the prediction of a company's CPM maturity through evaluating the BI development. It also shows if a high development of the BI items leads to a high development of the CPM items. In detail, the BI items most important to a successful CPM can be identified. Conversely, it should be possible to derive the BI tools and

processes most important to the effective use of a certain CPM process. This supports, in turn, the definition of the business value of BI.

## V. CONSLUSION AND FUTURE WORK

The Data Mining methods association rule discovery, clustering and decision tree modelling are seen as powerful research tools for determining the business value of BI on the effectiveness of CPM. Unlike the commonly used research methods like explanatory factor analysis and PLS, the Data Mining techniques include all research criteria, which is seen to give a more detailed insight and highly interesting new findings on the subject. After collecting the data sample, the above mentioned Data Mining methods will be applied. As indicated in Figure 1, the evaluation and discussion of the results will follow as a next step of this research. It is assumed that the findings will promote the clarification of the relationship between BI and CPM in more detail. In a future research, a time-lagged investigation is sought to evaluate the study results.

## REFERENCES

[1] M. Hannula and V. Pirttimäki, "Business Intelligence Empirical Study on the top 50 Finnish Companies", Journal of American Academy of Business, vol. 2, no. 2, 2003, pp. 593-599.

[2] M. Aho, "The Distinction between Business Intelligence and Corporate Performance Management-A Literature Study Combined with Empirical Findings", in Proceedings of the MCSP 2010 conference, 2010.

[3] O. Jacob and F. H. Lien Mbep, "Factors to determine the value of Business Intelligence to Corporate Performance Management", Hochschule Neu-Ulm, unpublished.

[4] Pro-How Paper 2/00, "Measuring the benefits of Business Intelligence", Available from http://legacy.wlu.ca/documents/22449/07_Measuring_the_Be nefits_of_BI_Viva.pdf, retrieved 2015.02.12.

[5] S. Negash, "Business Intelligence", The Communications of the Association for Information Systems, vol. 13, no. 1, 2004, pp.177-195.

[6] S. Williams and N. Williams, "The Business Value of Business Intelligence", Business Intelligence Journal, vol. 8, 2003, pp. 30–39.

[7] I. B. Pugna, F. Albescu, and D. Babeanu, "The Role of Business Intelligence in Business Performance Management", Annals of Faculty of Economics, University of Oradea, vol. 4, 2009, pp. 1025–1029.

[8] ResearchandMarkets, "Business Intelligence: Corporate Performance Management", Available from http://www.researchandmarkets.com/reports/1055897, retrieved 2015.02.19.

[9] J. Becker, D. Maßing, and C. Janiesch, „An Evolutionary Process Modell for the Introduction of Corporate Performance Management Systems", Data Warehousing, 2006, pp. 247–62.

[10] F. H. Lien Mbep, O. Jacob and L. Fourie, "Critical Success Factors of Corporate Performance Management (CPM)", BUSTECH, 2015, pp. 6-14..

[11] G. Richard, W. Yeoh, A. Y. Loong Chong, and A. Popovič, "An empirical study of Business Intelligence impact on Corporate Performance Management", Proceedings of the PACIS 2014 conference, 2014, Paper 341.

[12] S. Miranda, "Beyond BI: Benefiting from Corporate Performance Management Solutions", Financial Executive, vol. 2, 2004, pp. 58-61.

[13] N. Yogev, L. Fink, and A. Even, "How Business Intelligence Creates Value", ECIS 2012 Proceedings, 2012, Paper 84.

[14] M. L. Gargano and B. G. Raggad, "Data Mining – a powerful information creating tool", OCLC Systems and Services, vol. 15, no. 2, 1999, pp. 81-90.

[15] A. Stolzer and C. Halford, "Data Mining Methods Applied to Flight Operations Quality Assurance Data: A Comparison to Standard Statistical Methods", Journal of Air Transportation, vol. 12, no. 1, 2007, pp. 6-24.

[16] J. Jackson, "Data Mining; A Conceptual Overview", Communications of the Association for Information Systems, vol. 8, 2002, article 19.

[17] S. Natek and M. Zwilling, "Data Mining for small student dataset", Management, Knowledge and Leanring Conference, 2013, pp.1379-1389.

[18] J. Cleve and U. Lämmel, Data Mining, Oldenbourg Wissenschaftsverlag GmbH, 2014.

[19] H. Peterson, Data Mining: Methods, Processes, Application Architectures, Odlenbourg Wissenschaftsverlag GmbH, 2005.

[20] K. Backhaus, B. Erichson, W. Plinke, and R. Weiber, Multivariate Analysis – an application-oriented introduction, Springer-Verlag Berlin Heidelberg, Aufl. 13, 2011.

[21] B. Liu, Web data mining exploring hyperlinks, contents, and usage data, Springer-Verlag Berling Heidelberg, 2011.