Communication between Data Analysts and Domain Experts -

A Qualitative Case Study

Masterarbeit

Zur Erlangung des akademischen Grades „Master of Science (M. Sc.)" im Studiengang Wirtschaftswissenschaft der Wirtschaftswissenschaftlichen Fakultät der Leibniz Universität Hannover

vorgelegt von

Name: Buske   Vorname: Nicolas

█████   ████████   ████   ████████

Prüfer: Prof. Dr. M. H. Breitner

Hannover, den 28.09.2018

CONTENTS

1. INTRODUCTION

1.1. MOTIVATION

"Data Scientist: The Sexiest Job of the 21th Century."

– Davenport & Patil (2012b).

Today's business world is overflowed with Buzzwords like Big Data, Data Analytics or Data Science. Data-driven businesses emerge with growing extent and at a velocity, necessary structural changes in organizations cannot compete with. Recent megatrends like industry 4.0 and digitalization in general build the foundation for an innovative dealing with data. As an emerging part within organizations, Data Science provides opportunities for enhanced decision-making processes. As the success of businesses correlates predominantly with the quality of management decisions, those decisions are forced to be elaborated at best. Next to domain knowledge, experience and managerial intuition, Data Analytics supports the process of decision-making with growing extent (Provost & Fawcett, 2013a).

Hence, the processing and goal-oriented implementation of data emerged to one of the most relevant opportunities and challenges for organizations on a global scale. Referring to Davenport and Patil (2012b), Data Science enables organizations to gather more significant insights about their business and the process of decision-making will face disruptive alterations due to additional influencing factors in consequence of enhanced application of Data Analytics.

Yet, literature and research show that the correlation between domain knowledge and Data Science lacks investigation in respect of both qualitative and quantitative assertions. Next to possible influences of domain knowledge on Data Science, communication between Domain Experts and Data Scientists in general tends to have impact on Data Analytics. Thus, this thesis aims to find evidence whether the communication between Data Science and Domain Experts can influence the results of Data Analytics and how this communication is established in general. In addition, the impact of domain knowledge on Data Science will be investigated to identify possible flaws and mistakes within the process of interdisciplinary communication. To pursue studies in the above-stated explorative field, the qualitative method of expert interviews within the scope of various case studies is conducted.

Chapter two will introduce basic theoretical assumptions and frameworks to enhance the process of hypotheses building and to further discuss the introduced research question. Following, chapter three will contain the methodological approach used for

the expert interviews as well as a description of the case study research design. Chapter four then summarizes the main findings of the interviews in respect of the different cases and will discuss the findings in concern of the theoretical basics defined in chapter two. Moreover, chapter four will consider limitations of the research approach as well as practical implications for organizations and further research. Lastly, chapter five will conclude the research question and summarize the findings from literature and case studies.

1.2. PROCESS OF WORK

Considering the relevance and motivation for the research topic presented in the previous chapter, the process of this work is aligned to a step-by-step model by Mayring (2014). Within the first step the concrete research question is defined. Referring to the ongoing process of investigation, the research question will be linked to the outcomes of the extensive literature review by emphasizing relevant theoretical approaches and state of the art literature. On the one hand, domain knowledge of Data Scientists is seen as a key competence next to pure analytical skills, even if the expertise within different domains is limited. On the other hand, it is difficult to understand to which certain degree domain expertise has influence on the results of Data Analytics and in conclusion on decision-making processes and its results (Waller & Fawcett, 2013a). However, there may be a lack of knowledge integration on the site of Data Analysts as well as on the counterpart of the Domain Experts. To improve the relationship between those parties, clear and content-related communication is one of the main characteristics referring to valid decision-making processes. To be able to analyze the research topic of this thesis in a structured way, the following research question has been designed:

How does domain knowledge and communication between Data Scientists and Domain Experts influence Data Analytic processes?

Based on the research question it is possible to answer basic appreciations in respect of intraorganizational communication and the influence of domain knowledge from a general point of view. This will be part of the theoretical introduction, which takes place in chapter two and introduces different insights to the reader about what Data Science and Domain Expertise is about at its basics. Moreover, it is crucial to provide the reader with a brief understanding of intraorganizational communication and how this can

possibly influence the results of data analyses, to ensure a better recognition about the qualitative research conducted in the second part of this thesis. Within this chapter common sense is created in respect of the relevant topics, definitions and used terms within this thesis.

The third, fourth and fifth steps referring to Mayring (2014) include the introduction of the research design and methodology. This topic is part of chapter three, by presenting the research approach followed by the procedure within the case-study design. To enhance the comprehensibility for the reader, a definition of the sample regarding the expert interviews is provided. The presentation and analyzation of the results is defined by the sixth step of the process and referred to in chapter 4 of this thesis. In this chapter, the initial research question is answered in respect to the gained results by comparing the different cases as well as the results of the literature review. The step-by-step model is completed with the discussion of the results and can be summarized by the following Figure:
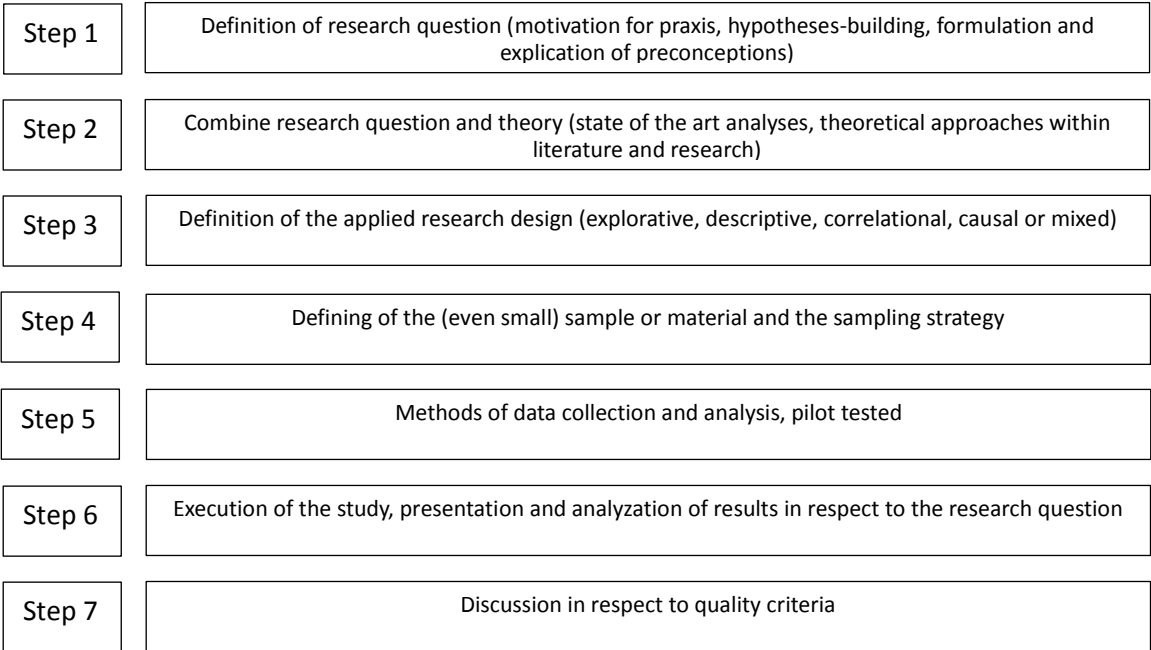
Step 1	Definition of research question (motivation for praxis, hypotheses-building, formulation and explication of preconceptions)
Step 2	Combine research question and theory (state of the art analyses, theoretical approaches within literature and research)
Step 3	Definition of the applied research design (explorative, descriptive, correlational, causal or mixed)
Step 4	Defining of the (even small) sample or material and the sampling strategy
Step 5	Methods of data collection and analysis, pilot tested
Step 6	Execution of the study, presentation and analyzation of results in respect to the research question
Step 7	Discussion in respect to quality criteria

Figure 1 Step by Step Model Source: Own Representation based on Mayring, 2014

5. CONCLUSION

This thesis introduced the question how domain knowledge and communication between Data Scientists and Domain Experts influences Data Analytics. To find proof on this, chapter two introduced the recent theoretical approaches on the topics Data Science, expert knowledge and intraorganizational communication. Building on that theory, this thesis introduced four main hypotheses. To test these hypotheses, expert interviews were conducted within four practical cases. In advance of the research process, conceptual frameworks were introduced. In the following, the findings were presented and discussed in respect of the single cases as well as regarding similarities and differences concerning literature. Lastly, this thesis introduced both limitations and practical implications for further research.

The main findings of this paper are that domain knowledge has a theoretical impact on Data Science, that cannot be confirmed to the same degree in practice. Although the practitioners in the investigated cases agree on the enhanced recognition of Data Science and positive impacts of domain knowledge from a theoretical point of view, appropriate behavior in dealing with domain knowledge and the acquisition of domain knowledge is almost unattended in practice. Furthermore, domain knowledge in correlation with communication in Data Analytic processes can enhance the motivation and performance of Data Scientists. Moreover, the process of communication in collaborative projects enables Data Scientists to overcome the lack of relevant domain knowledge to a certain degree to significantly enhance process and outcomes of Data Analytics.

The results and findings of this thesis can be of interest for other researchers in the field of Data Science as well as for practitioners to adjust their dealing with domain knowledge in respect of Data Science and thus, to enhance processes and outcomes of Data Analytics and concluding the overall performance of organizations. Regarding further researchers, this thesis offers new insights on the correlation between Data Science and domain knowledge and can hence be used as fundamental research to generate deeper understanding on the explorative overall topic.

Due to the nature of qualitative research, the findings of this thesis should be seen as orientation for further investigations and not as facts. Nevertheless, this study can be used to find more specific evidence in the field and to quantify the qualitative results.

Closing, one can conclude that evidence is found that domain knowledge of Data Scientists and the process of communication between Data Scientists and Domain Experts have impact on the process and results of Data Analytics.