Strategic Approaches to the Use of Data Science in SMEs

Lessons Learned from a Regional Multi-Case Study

Cornelia Ferner and Thomas Heistracher
Salzburg University of Applied Sciences
Salzburg, Austria
firstname.lastname@fh-salzburg.ac.at

***Abstract*—The potential of data analytics and modeling, especially in the context of digital transformation, is well known and seen as an opportunity to increase competitiveness and innovation in companies. However, the technical and organizational implementation of these methods poses challenges specifically for SMEs. A regional multi-case study explores the causes and exhibits patterns that suggest differentiated recommendations for action. Awareness of the diverse limitations and challenges can help all players involved – companies, universities and governmental organizations – to design future projects in an even more targeted manner.**

***Index Terms*—data science, digital transformation, SME, triple helix**

I. Introduction

Data science is a team effort that usually requires several roles: A data scientist that is not only proficient in mathematics and programming, but can also apply machine learning methods to target applications. A data engineer, who is responsible for the physical data storage, which includes infrastructure design, data access, data security etc. A data analyst that analyzes existing data and creates visualizations, reports or dashboards. Higher-level management tasks are usually taken over by a Chief Data Officer (or Data Manager), who reports to the executive management [1].

While big companies have increasingly large data science teams, SMEs most often do not have comparable capacities. Fortunately, this does not need to keep them from developing innovative ideas and implementing data-centric projects. Innovation and economic development is often driven by the successful collaboration of three players: academia (universities), governmental organizations and the industry. This interplay is known as Triple Helix Model [2], that formalizes the interdependencies and interactions between the three players responsible for a region's long-term technological development.

Exchange with and input from universities and other research facilities can help with the realization of new ideas. University-industry collaborations provide access to additional know-how and human resources outside the company and often enable testing of new ideas without risk to the production environment. Companies cannot flourish without the infrastructure and legal framework provided by the government. Besides legal frameworks for copyrights, taxation and location policy, the government's role is to set funding programs to encourage collaboration and to provide incentives for innovation. The interplay between government and academia defines the long-term strategic research direction and can shape both the academic landscape as well as the job market.

The federal state of Salzburg offers a specific funding program designed to support the digitization efforts of regional companies, mainly SMEs. In order to assess the program, the currently funded projects in the context of data science were evaluated. The companies that participated in the multi-case study were selected with the support of the regional innovation service agency, taking care to cover as diverse a spectrum of industries, company sizes within the category SME and geographic locations within the federal state of Salzburg as possible; the ICT sector and e-commerce businesses were deliberately excluded here.

As a result, twelve semi-structured interviews [3] were conducted via video calls with experts from the companies who were directly involved in the project implementation or who played a leading role in the strategic development (e.g. owners, (technical) managers etc.). The guideline for the interviews included closed questions that are in line with the levels and dimensions of the Data Science Maturity Model (see Section II) and qualitative questions to capture the companies' self-assessment from their own perspectives, especially regarding risks and future opportunities that came from their projects.

The contribution of this paper is two-fold: First, we present the findings of our multi-case study, where we analyze regional SMEs based on a simplified data science maturity model and highlight interesting insights. Second, we derive

The study presented in this publication was conducted as part of the research project "DataKMU – Vernetzung und Wissenstransfer im Bereich Data Science" that is funded by the European Regional Development Fund with grant number AB215.

recommendations for the three categroies of players involved to even increase future project output and foster innovative collaboration.

II. Data Science Maturity Model

The Data Science Maturity Model [4], introduced by Mark Hornick, Senior Director at Oracle Data Science and Machine Learning, allows to assess a company's current state concerning data science strategy and technology:

"Enterprises that already embrace data science as a core competency, as well as those just getting started, often seek a roadmap for improving that competency. A data science maturity model is one way of assessing an enterprise and guiding the quest for data science nirvana. Upping an enterprise's level of data science maturity enables extracting greater value from data for making better data-driven decisions, realizing business objectives more efficiently, and having a more agile response to changing market conditions." [5], p.3.

The data science maturity model identifies ten dimensions that are decisive for the success of data science projects. Each dimension consists of five maturity levels - from basic level 1 to the most mature level 5. Each company can individually specify at which level they are or intend to be in each dimension.

Given that the surveyed companies do not have dedicated data science teams, their profiles where assessed in five dimensions: *strategy*, *data management*, *methodology*, *tools* and *deployment*. The dimension *strategy* captures the strategic orientation of a company with regard to data use and exploitation and assesses, whether data is regarded as a by-product or as capital. The dimension *data management* captures the type of data storage that is used – central or distributed systems locally in the company or outsourced to external providers. The *methodology* dimension captures how companies use methods to analyze the past and forecast the future. Within the dimension *tools*, the usage and scalability of software packages is monitored. The fifth dimension, *deployment*, describes how results are reported, from static files to dashboards to continuous deployment of dynamic models.

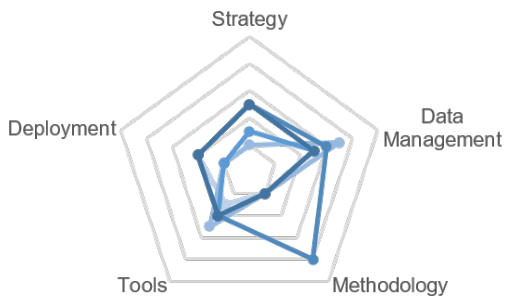
III. Multi-Case Study: Data Science in SMEs in Salzburg

Even if not every company needs to develop a data-centric business model [6], the current focus is on driving digitization, at least in certain company areas. Recording and using data alone is not sufficient. The focus needs to shift towards exploiting the data value by creating new insights and deriving action items. In this context, we talk about data science [7], i.e. the analysis and modeling of data.

In our multi-case study, the experiences and findings of twelve SMEs in Salzburg concerning their implementation of data science projects were surveyed. In the absence of in-house IT and software development teams, let alone dedicated teams for data science, digitization projects in these companies are fundamentally collaborative efforts. Only if

(a) *Pioneers*' company profiles

(b) *Strategists*' company profiles



(c) *Pragmatists*' company profiles

Fig. 1: Different company profiles with regard to the five dimensions of the Data Science Maturity Model. The three groups were assigned through clustering. The different color intensity is used to discriminate the individual company profiles.

Pioneers	
Pioneers	Legal questions with respect to machine learning
	Infrastructure, equipment
	Acquisition costs
	Application-specific challenges

(a) *Pioneers*' challenges

Strategists	
Strategists	Missing manpower and/or lack of specific training
	Cost-benefit analysis
	Costly data management
	Extensive data maintenance

(b) *Strategists*' challenges

Pragmatists	
Pragmatists	Data transparency and security
	Low data volumes ("small data")
	Lacking interfaces to current software
	Lacking overview of methodologies and available software tools

(c) *Pragmatists*' challenges

TABLE I: Main challenges in data science projects mentioned by the three different company groups.

motivated employees get involved beyond their actual duties and are open to new ideas such projects can be realized at all.

When asked about the motives for implementing data analytics projects, intrinsic motives predominated in the interviews. The desire for further development and technological progress, the expansion of the product portfolio or services offered, as well as resource optimization (time, costs) were named as reasons. In addition, legal and certification requirements or explicit customer wishes are also drivers of digitization.

Systematic analysis of individual company profiles with regard to the five data science maturity dimensions reveals patterns, as shown in Figure 1: A group of companies, which we refer to as the *Pioneers*, achieve the highest level of proficiency in some dimensions and act in a data-centric manner with regard to strategy, management and implementation (see Figure 1a). Another group of companies shows a peak at the *strategy* dimension (see Figure 1b), indicating that these companies are aware of the value of their data and the benefits they can derive from it. We refer to this group as *Strategists*. The profiles of the last group of companies does not show this peak (see Figure 1c). The strategic use of data plays a less important role for them than the actual implementation. This third group is referred to as *Pragmatists*.

The clear differences between the company profiles that allows a grouping into three distinct clusters rises the question, which underlying factor is common to the companies within each group? Which factors determine the different approaches towards data science projects? Reasons such as company size or culture, or a common industry sector can be ruled out, as those differ within each group anyway.

An explanation for the different maturity profiles is revealed by the answers the companies provided to the question about the main challenges when implementing data science projects (see Table I). While those challenges are very application-oriented among the *Pioneers* (see Table Ia), decisive differences can be identified between the *Strategists* and *Pragmatists*:

The *Strategists* are most likely to see the availability of know-how and trained employees as a challenge (see Table Ib). With the implementation of data-driven projects, the demands with regard to employees would increase, which would require different training or further education. In addition, companies from the *Strategists* group mention the difficulty of recruiting new employees with the appropriate training, as salaries comparable to the IT industry would often exceed their own salary structure. The *Strategists* also draw attention to the need to weigh up the costs and benefits before implementing data-driven projects, which also underscores the clear peak in the strategy dimension. The cost-benefit consideration remains an issue even after successful implementation, when it comes to the effort required for continuous data management and maintenance.

The *Pragmatists*, on the other hand, cite legal considerations such as data transparency and security as a fundamental challenge. When implementing data science projects, they find the provision of interfaces and subsequently the connection of data science tools to existing software to be challenging. In this context, the lack of comparability and confusion of available tools is also addressed (see Table Ic).

The differences between the *Pragmatists* and the *Strategists* result from the type of data the respective companies work with. Companies with a clear strategy for how to use the data own that required data. They are aware of the available volume and quality and can make an initial assessment of what information can be generated. Companies in the *Pragmatists* group are largely dependent on the use of external data they do not own themselves. The availability or possibility of use as well as the data quality are comparatively uncertain or unknown. Table II lists examples of different types of data according to their source and the volume to be expected.

IV. Results

Data Science and Machine Learning methods that are applicable in a project depend on the type of data and the available volume. In addition to fundamental data protection issues, the primary challenge especially in the case of external data is the merging of heterogeneous sources. Missing interfaces and a common data format need to be implemented. In some cases, data must be made accessible step by step from external sources or alternatively generated by suitable methods within the company itself. This problem is referred to as cold start problem: Many methods for data science require large amounts of data to be successful in automated analysis and modeling. If this data is not (yet) available in sufficient quantity, simpler methods can be applied in the short term or data can be simulated as a workaround.

Volume	Source: internal	Source: external
high	Plant data Machine data Control signals Measuring signals Process data Log data	Network data Access data Utilization data Log data Energy consumption data
low	Employee data Customer data Order data	Data from customers Data from suppliers Data about products, parts Data about (raw) materials

TABLE II: Matrix with different types of data according to their source and volume to be expected.

A different challenge is posed by a high number of internal data sources: Machine and process data in particular are often available in very high resolution (e.g. every millisecond) and thus in large quantities. This raises the question of suitable data reduction methods either by pre-filtering or aggregation. Applicable methods mainly depend on the amount of data to be processed or on the type of interfaces required.

For each project and each task, it is therefore necessary to assess the ”data inventory” and its requirements in order to be able to develop customized solutions. A pure blueprint implementation is only possible in the rarest of cases. This is why domain expertise with extensive knowledge of the (business) processes that generate the data is needed, as well as data science experts with the skills to select suitable methods and adapt them to the specific needs [8].

From these insights, we derive recommendations for each of the players of the triple helix:

A. SMEs

Companies can assess their current data science maturity level based on the model presented in Section II. The resulting profile allows to align the future strategy and goals. In addition to the maturity profile, a ”data inventory” creates awareness of internally available data and possible external dependencies, which sets the technological and methodological frame for future projects. As more and more jobs require working with data, it is recommended to promote data literacy among all employees in the company.

B. Universities

Data science projects are optimal starting points for industry-university collaborations: SMEs with limited trained staff in programming and/or data science can benefit from the external technological know-how of academic reasearchers and can contribute the required domain expertise to scientific research projects. In addition, collaborations with universities provide access to an extended pool of personnel.

From a university's point of view, a data science project allows for interdisciplinary collaboration, as applied data scientists can team up with business informatic scientists and machine learning researchers.

C. Government

With their digitization program, the government of the federal state of Salzburg has filled a funding gap by especially targeting regional SMEs that take their first steps towards more digitization. While many companies have already benefited, the results from our structured interviews show another potential: An additional benefit can arise from extending the programs to also grant funds to companies owning data, but not actively developing data science projects. The provision of data (e.g. by suppliers, manufacturers, producers) can help other companies to get access to ”missing links” in their projects through ”data partnerships” [8]. Allowing such companies to jointly apply for funding would help in making more data (publicly) available and framing data as a business case. This could also serve to bring about a cultural change towards open data in the realm of data science.

V. Conclusion

As part of a survey of selected companies in the province of Salzburg, the importance of the topics of digitization and data science was surveyed, especially for SMEs in non-ICT industries. All companies recognize data science as an added value for their business models and have gained initial experience in implementing data science projects. The evaluation of the interviews shows a differentiated picture regarding the companies' approach to these projects that is partly caused by the availability and volume of the required data. As the appropriate methods and tools for generating data insights and predictions depend largely on the type of data, smaller explorative studies can help in evaluating approaches and assessing their effectiveness systematically without risk. By applying for advanced funding programs, university partners can also be involved in research projects that step-by-step lead to higher data science maturity levels.

References

[1] T. Stobierski, “How to Structure Your Data Analytics Team,” *Harvard Business School Online*, 2021, available online: https://tinyurl.com/rtmaxvk.

[2] H. Etzkowitz, *The Triple Helix: University-Industry-Government Innovation in Action*. New York, NY: Routledge, 2008.

[3] A. Bayman and E. Bell, *Business Research Methods*. Oxford: Oxford University Press, 2011.

[4] M. Hornick, “Data Science Maturity Model - Summary Table for Enterprise Assessment,” Oracle R Technologies Blog, 2018, available online: https://blogs.oracle.com/machinelearning/post/data-science-maturity-model-summary-table-for-enterprise-assessment-part-12.

[5] M. Hornick, “Data Science Maturity Model for Enterprise Assessment,” Oracle, Tech. Rep., 2020, available online: https://www.oracle.com/a/devo/docs/data-science-maturity-model.pdf.

[6] S. Andriole, “Five Myths about Digital Transformation,” *MIT Sloan Management Review*, vol. 3, no. 58, pp. 20–22, 2017.

[7] J. Kelleher and B. Tierney, *Data Science*. Cambridge: The MIT Press, 2018.

[8] U. Kruhse-Lehtonen and D. Hofmann, “How to Define and Execute Your Data and AI Strategy,” *Harvard Data Science Review*, 7 2020, available online: https://hdsr.mitpress.mit.edu/pub/4vlrf0x2.