NOVA IMS
Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

Designing a collaborative AutoML tool to help organizations become data-driven

Jens Mattmüller

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management with a specialization in Information Systems and Technology Management

Advisor: Prof. Vítor Santos, PhD

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

ABSTRACT

This study aims to address a lack of knowledge in the emerging field of automated machine learning (AutoML) techniques. While the AutoML technology develops further and further and provides increasingly robust and interesting results, there is only little to no current research on how this technology can be adopted and scaled across different functions and teams of any organization. Thus, this study raises the research question of how an information system that leverages AutoML techniques can empower organizations and their non-technical individuals to collaborate on and adopt machine learning techniques in their daily lives to unlock the value of available data. To gain a clear analytical lens, this study is conducted in the environment of Management Consulting Companies (MCCs) as they span all industries and multiple tasks within diverse organizations and therefore promise a good transfer of knowledge to other application areas. A special emphasis is given to non-technical users and the possibilities of them participating in such a system as that has the potential to reach a large number of real-world practitioners. The identified problem is tackled with a Design Science Research (DSR) approach. A workflow of how an information system can support its users to leverage AutoML serves as an artifact that is evaluated by experts. Learnings from the theory behind the proposal and its evaluation contribute to literature around AutoML and the transformation of the MCC industry as well as practical applications in both fields. Results suggest that AutoML is best used to conduct quick experiments and find out which applications have the highest business value before involving experts. Major challenges are to help non-technical users define a use case and prepare data.

KEYWORDS

#AutoML; #collaboration, #Data-driven decision making; #Machine Learning; #DesignScienceResearch

Index

List of figures

List of Tables

List of Abbreviations and Acronyms

AI	Aritificial Intelligence
AutoML	Automated Machine Learning
MC	Management Consulting
MCC	Mangement Consulting Companies
ML	Machine Learning
MLOps	Machine Learning Operations
NAS	Neural Architecture Search
PoC	Proof of Concept

1. INTRODUCTION

1.1. BACKGROUND AND PROBLEM IDENTIFICATION

Machine Learning research and business applications have seen a massive rise in popularity throughout the last decades (Borges, Laurindo, Spínola, Goncalves and Mattos 2021). Recently, an interesting sub-field to the big machine learning conglomerate has developed: automated machine learning (AutoML). A good definition of AutoML is given by Hutter, Kotthoff and Vanschoren: "The field of automated machine learning (AutoML) aims to make […] decisions in a data-driven, objective, and automated way: the user simply provides data, and the AutoML system automatically determines the approach that performs best for this particular application" (2019, p. 9). The big spark of hope here is to make machine learning techniques no longer available only to cases where programmers with specialized skills are available. This is particularly desirable as there is a shortage of skilled workforce able to conduct (big) data analytics tasks (Asamoah, Sharda, Zadeh and Kalgotra 2017).

Within the science of AutoML, the so called "human in the loop" narrative swapped over from general Artificial Intelligence (AI) research. This field of studies emphasizes how humans interact with intelligent systems to achieve the best results (Seeber et al 2020 and Ostheimer, Chowdhury and Iqbal 2021). But mostly, these researchers focus on experts intervening with the system and do not take a broad organizational perspective on the problem (i.e.: how do different roles within an organization work with AutoML techniques?). Generally, in the field of advanced data analysis techniques (such as machine learning) researchers seem to either focus on capability-related challenges of making advanced analytics or machine learning techniques work (e.g. Schüritz et al. 2017, Bose 2009, Gupta & George 2016) or on dominantly technical aspects of certain use cases (e.g. Kitchens et al. 2018, Yuheng et al. 2019). On the other hand, Prof. Ron S. Kennett and Prof. Thomas C. Redman point out in their book "The real work of Data Science" (2019) that the biggest challenge in becoming a data-driven organization is not just applying the fanciest algorithms but more about where relevant skills are located in an organization and how they are scaled across different teams and their business needs. There is a lack of research that combines organizational and technical perspectives to answer the right questions of how organizations can become data-driven: how are non-technical individuals able to identify use cases within their organizational context? How do they collaborate on and apply machine learning models as a process? How do business experts and data analysts work hand in hand to ensure that technology aligns best with business needs? For a whole organization to become data-driven, non-technical individuals must be included in scaling AI and Machine Learning techniques along their business problems. Therefore, I propose the research question:

How can an Information System (IS) that leverages AutoML techniques empower organizations and their non-technical individuals to collaborate on and adopt machine learning techniques in their daily lives to unlock the value of available data?

I find this perspective particularly compelling as providing non-technical users with the abilities to leverage machine learning techniques has the potential to scale the participation in creating value from data across every organization. Managing the scope of this thesis and trying to find a clear analytical perspective to base my conclusions on, I decided to give special emphasis on Management

Consulting companies for the design of said information system. MCCs are especially eligible to conduct this analysis on as they work for every industry all over the world and are seen as agents of change concerning the adoption of innovative technologies (Curuksu 2018). Therefore, designing the tool based on MCCs needs promises to allow for the design of an industry-agnostic tool that could be adopted quickly in any other organization.

Following a Design Science Research (DSR) approach, the succeeding chapters consist of an analysis of the existing knowledge base and the context of the problem at hand. Synthesizing the learnings from before, a proposal is being developed. After evaluation, this proposal will be discussed under the light of existing and emerging concepts in research. Drawing a conclusion, I will also mention limitations and future research directions for the adoption of AutoML-enabled tools in organizations.

1.2. Study objectives

The research objective is to design an AutoML tool that allows non-technical users to execute and collaborate on machine learning projects in order to scale data driven value-creation across organizations with a special emphasis on Management Consulting companies.

To achieve this goal, the following intermediate objectives are defined:

1. Diagnosis of companies needs with a focus on MCCs
2. Identification of the areas/ tasks suitable for applying AutoML tools
3. Proposal of a workflow that shows how an AutoML tool can help organizations use machine learning techniques
4. Creation of a demonstration for the proposed workflow
5. Evaluation and ideas for improvement of the proposal

1.3. Study Relevance and importance

From a practical point of view, the importance of the study becomes clear if we look at recent economic developments, here on the example of just one industrialized country: The German Chancellery estimates that the value creation potential of the data economy – just in Germany - will amount to over 400 billion euros by 2025 (2021). However, according to the same study, in Germany alone over 90 percent of the data available in organizations has not yet been used (German Chancellery, 2021). Another study by the Federation of German Industries (BDI) shows that only 23 % of their interviewed companies have a regular strategic process in place to scout use cases and sources for data usage and that 45 % of the interviewed companies don't use data to optimize their products and business models at all, yet (BDI, 2021). I have made similar experiences in my professional life: a lot of organizations understand data as a critical resource and say they want to work data-driven but fail to incorporate data analytics in their daily lives and decision making. On the other hand, I have seen the variety of use cases and how powerful advanced machine learning techniques can be when implemented in the right business context, which motivates me to conduct this thesis. The gap between potential of value-creation with data and its reality could (partly) be bridged by an easy to use, collaborative AutoML tool that helps organizations apply advanced techniques like machine learning to their data more efficiently.

Focusing on MCCs in this study provides an analytical lens that enables researchers in the future to transfer the achieved insights faster than with any other industry as management consulting is applied in each domain and problem environment. On the other hand, it is a very interesting point in time to look at the management consulting industry as the industry is in the process of transforming their traditional business model (Christensen, Wang and van Bever 2013 and Curusku 2018) and there is only little research on the role of advanced analytics techniques like machine learning in this process.

Through the experimental character of this research (Design Science Research approach), the study will contribute to practical knowledge about how an Information System leveraging AutoML would have to be designed to have organizations (and especially MCCs) adopt it. Designers, software developers and CIOs can benefit from these findings.

On the academic side, I have already discussed the research gap in the field of human interaction and organizational adoption of AutoML systems in the introduction. This study will contribute to the advancement in knowledge of how AutoML systems have to be designed to be adapted by non-technical users and about the potentials of collaborating on AutoML systems. Further, it will contribute to the research on business model transformation in the management consulting industry.

In March and Smith's famous matrix (1995) the contributions of this thesis can be seen in the construction of a model to understand why non-technical users use advanced analytics methods and in building and evaluating an instantiation.

2. METHODOLOGY

As Brock, Hevner and Maedche (2020) state, Design Science Research (DSR) is "a problem-solving paradigm that seeks to enhance human knowledge via the creation of innovative artifacts" (p.1). Knowledge is created in the process of designing a new artifact and in testing and evaluating if and how it enhances the problem context where it is applied. Therefore, DSR provides a very valuable approach for the research question at hand. On the one hand, it provides a structured approach to designing and testing the collaborative AutoML tool. On the other hand, its experimental nature allows to iteratively develop a solution with predefined evaluation steps, which enhances the practical feasibility and probability of real-world adoption.

2.1. DESIGN SCIENCE RESEARCH

Peffers et al. (2008) summarized the typical DSR activities in the following graphic:

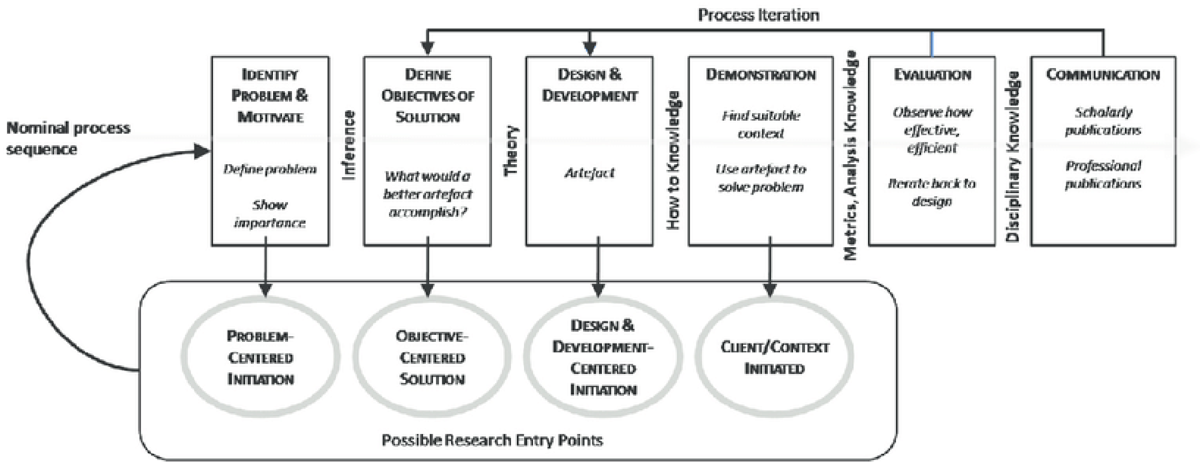


Figure 1: Design Science Research according to Peffers et al. (2008)

After the definition of a problem and the corresponding objectives for a solution, an artifact is created. The step of creating an artifact involves using an existing or creating a new theory of how an artifact would have to look like to appropriately solve the defined problem. After, the artifact is demonstrated within the problem context and tested for its ability to solve the given problem (e.g. in an experimental study, simulation or case study). An evaluation with the learnings of the demonstration step is executed and the artifact should be improved iteratively in this process. Lastly, the produced knowledge should be communicated.

2.2. RESEARCH STRATEGY

Applied to this master thesis, the problem identification is represented in the research question and the objectives have been mentioned in chapter 1.3. As there were no sufficient theoretical frameworks identified, which give a confident base to create an innovative artifact from, theorizing will be an important step in this master thesis to enable the design of an artifact. The following graphic by Hevner et al. (2004) gives another perspective on the DSR approach and emphasizes how building the artifact interacts with the existing knowledge base and the environment of the problem context. While the environment highlights the needs a solution has to serve, the knowledge base gives an understanding of existing foundations and methodologies that are available to apply to the problem.

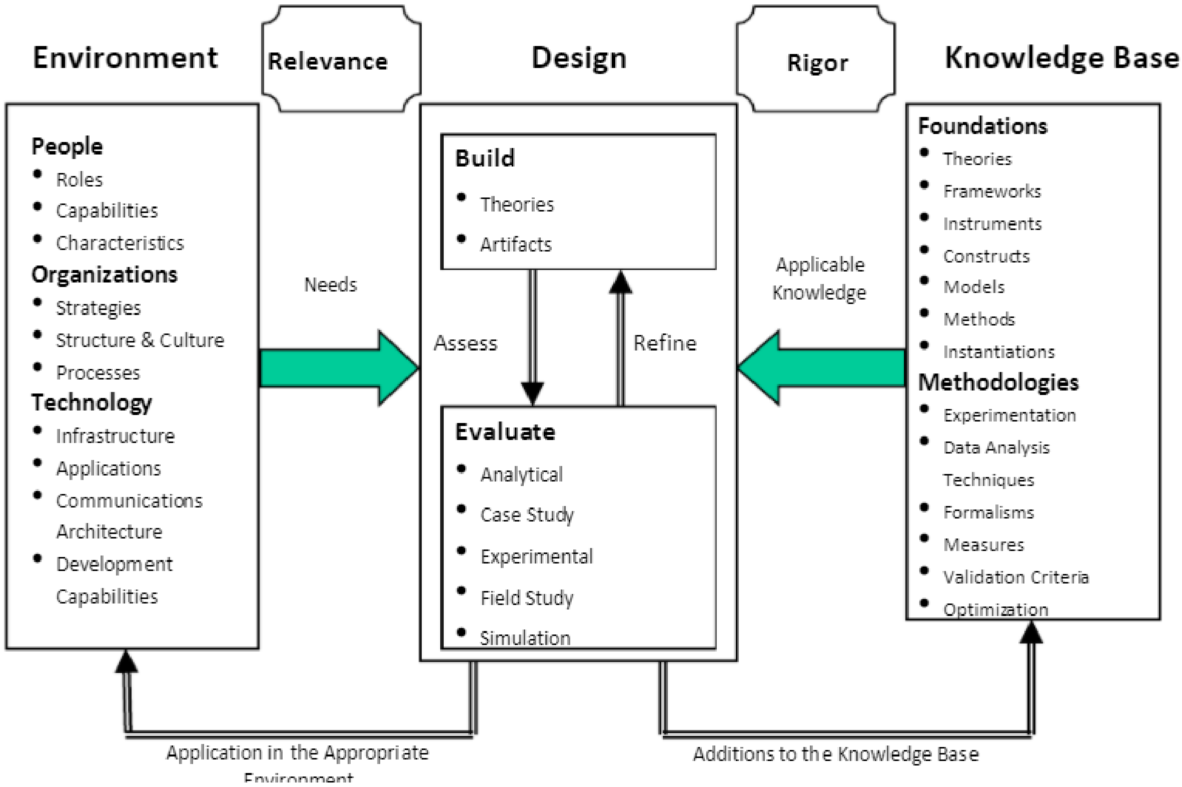


Figure 2: Different perspective on the Design Science Research Approach by Hevner et al. (2004)

In the context of this research effort, the environment is represented by the organizations and the people within them who shall be users of the tool that is to be created. Here, I will give special emphasis to MCCs to narrow down the applicable environment of the problem space as mentioned before.

After a structured literature review on the state of the Management Consulting industry and its current challenges, I will apply a SWOT analysis to summarize and pinpoint strengths, weaknesses, opportunities and threats to MCCs. For the structured literature review I analysed the biggest online databases for scientific literature (i.e. Scopus, Web of Science, IEEE Exploe, ScienceDirect) with the search terms: #consulting industry, #management consulting #consulting disruption, #consulting AutoML. I proceeded to assess the most relevant papers from these databases.

SWOT analysis is used in various ways in academic literature, for example to analyse optimal strategic positioning of companies or whole industries (Helms and Nixon 2010). Understanding the strategic positioning of MCCs with regards to digital disruption is key to create a proposal that is relevant to the scientific and practical audience.

On the other hand, the existing knowledge base on machine learning with a focus on AutoML and related trends constitutes the space in which we search for solutions to apply in the problem environment. Thus, a structured literature review on Machine Learning with a focus on AutoML is conducted where I analysed the biggest online databases for scientific literature (i.e. Scopus, Web of Science, IEEE Exploe, ScienceDirect) with the search terms: #Machine Learning, #AutoML #Humans in the Loop, #MLOps, # AutoML and enterprises. I proceeded to assess the most relevant papers from these databases.

Understanding the problem environment and the existing knowledge base, I am able to enter the Design phase with building a theory on the most important characteristics for a tool to empower organizations and their non-technical individuals to collaborate on and leverage AutoML techniques in their daily lives. The theory will be expressed in high-level assumptions which act as a frame for further exploring the expected behaviour for a solution in the defined problem environment. Thus, the high-level assumptions enable me to describe a scenario in which individuals within a MCC (collaboratively) use AutoML to create value for a client organization (Use Case). As a textual description of a Use Case cannot be seen as a testable artifact yet, it is necessary to further detail requirements for a solution. In Software Engineering it is a common technique to derive requirements from a textual description of a Use Case (Pressman 2009, p. 154 ff).

When describing a system to be developed requirements are usually separated into functional and non-functional requirements (Sommerville 2010). Functional requirements describe the behaviour of a system while non-functional requirements “apply to the system as a whole rather than individual system features or services” (Sommerville 2010, p. 85). While defining requirements one needs to have a good understanding of a possible user. Like Karmaker et al. (2021), I differentiate two different possible types of key users to an AutoML tool:

- “Domain Expert: A person who is fluent in the domain where ML is being applied, but has minimal knowledge of how ML itself works.
- Data Scientist: A person who knows how ML works but has minimal knowledge of the domain” (p. 2).

In the context of this study Domain Experts are non-technical individuals working in MCCs while Data Scientists could work within an MCC or for another company. I often use the term “business user” to refer to a Domain Expert and “ML expert” to refer to a Data Scientist. When talking about the Management Consulting industry it is usually the case that another party is involved: the organization that consumes services of MCCs (or even more specifically the organizations for whom a MCC might develop ML models). I refer to these organizations as “clients” or “client organization” in accordance with other scholars (compare for example Granzen et al. 2019 and Curusku 2018).

To bring together the perspective of users and activities that are performed and refine the Use Case described before, I use a Unified Modelling Language (UML) activity diagram. A UML activity diagram “depicts the dynamic behaviour of a system (…) through the flow of control between actions that the system performs” (Pressman 2009, p.853). This UML activity diagram will represent my artifact as it gives a clear idea of how the proposal can help if applied in the problem environment. Further, it can be evaluated and iterated upon better compared to a textual description which leaves more room for ambiguity. Adhering to an international best practice like UML for representing my proposal makes it equally understandable for academic scholars as well as practitioners.

The evaluation of the artifact will take place in three expert interviews. To demonstrate the proposal, I will build an example for a user interface from the UML activity diagram. The user interface is built in Microsoft Power Point and imitates a software product that demonstrates the desired functionalities of my proposal. After a demonstration of the proposal, all experts will be asked the same three questions:

1. What is your general impression of the proposal?

2. Do you think you/ your organization would use the proposal if it was a real product?

3. What are your recommendations to improve the proposal?

The selection of the experts is based on experience in the fields of machine learning and the Management Consulting industry.

3. LITERATURE REVIEW – MACHINE LEARNING

In this chapter we are going to explore the recent developments and business relevance of machine learning, which led to AutoML emerging as a promising technology. Doing that, we will also observe the adoption and practical appliances of machine learning and AutoML outside of academia. There are two reasons why combining these two perspectives makes sense: First, AutoML is a relatively young discipline in research, so there aren't many practical analyses on its adoption outside of academia available. Further, by looking at practical implementations of AutoML, we will better understand which factors lead to successful adoption of such tools. I will give a quick introduction to machine learning in general to start this chapter but cut through that part rather quickly as there is extensive literature out there in every direction and for the scope of this thesis there is enough to explore in AutoML and its business relevance.

3.1. INTRODUCTION: CONCEPTS AND APPLICATION AREAS OF MACHINE LEARNING

Machine Learning is a sub-field of AI and can broadly be categorized into supervised learning, unsupervised learning, and reinforcement learning (Maleki et al. 2020). The main difference is that while in supervised learning training data is labelled, it is not in unsupervised learning. One may also include semi-supervised learning into the mix, which works with a small amount of labelled training data and a larger amount of unlabelled data (Maleki et al. 2020). Supervised learning is often used in prediction efforts, such as the classification of which customers churn using algorithms like decision-trees or ensemble classifiers (De Bock and Caigny 2021). Unsupervised learning methods do not have a target variable to predict but try to find patterns in datasets, for example to cluster customers with algorithms like k-means (Li et al. 2021). Another class of unsupervised learning algorithms is dimensionality reduction such as Principal Component Analysis (PCA), which has a big meaning for working with huge datasets (Chen and Han 2021). Reinforcement learning in contrast is all about basing decisions on past experiences and therefore a stepwise class of algorithms, which have become popular by beating humans in complex tasks such as the game Go or Chess (Silver et al. 2018). But reinforcement learning also has various use cases in business, for example for decision making in finance (e.g. Song et al. 2021 & Olschewski, Diao and Rieskamp 2021).

With more and more practical applications of machine learning rising and an increased interest in literature, specifically neural networks are a part of machine learning that has gained significant traction in the last years (Samek et al. 2021). So called deep neural networks (neural networks with a complex structure of layers and neurons) do not only enable working with very complex tabular datasets but also are the fundament of advanced applications such as autonomous driving (Tseng, Lin, Chen and Hassan 2021) or speech and text recognition (Alshemali and Kalita 2020). Moreover, another field that has been trending within the last years are ensemble classifiers for everyday prediction task. The idea behind ensemble classifiers is intuitive: if one algorithm or machine learning model can be distorted or incorrect, why don't we apply a multitude of models and let them vote on a final output? Emerging already in the 1990s (Ho 2002), they by now have proven to be extremely efficient and usually outperform single classifiers that are manually tuned (Sagi and Rokach 2018).

Studying the enterprise adoption of machine learning techniques, Lee and Shin (2020) state that there are four major challenges for organizations today. While ethical and data-quality challenges as of now can only be tamed trough human intervention (compare chapter 3.2.1), the challenges of having not enough human resources to develop and deploy machine learning models as well as insecurities about cost-benefit ratios can be mitigated trough AutoML (compare chapter 3.2.2). This happens on the one hand through enabling more people to execute machine learning models and subsequently by dramatically lowering the hurdle to develop proof of concepts. Lee and Shin (2020) state that “managers need to have a clear understanding of its value-generation mechanism.” (p. 168) concerning the decision to implement a machine learning technique, which works best whenever a first prototype of a possible solution is drafted early. We will explore how AutoML transforms the organizational workings of machine learning adoption in the following chapters but will also take a look on how machine learning is operationalized in organizations trough combining development and operations practices with a concept called MLOps.

3.2. Machine Learning Operations (MLOps)

Although developing machine learning applications has become cheaper and faster over the course of the last years, there is a risk of falling into the pitfall of not considering the long-term costs of these applications. As Google-researchers Sculley et al (2015) point out there comes a hidden *technical debt* (a term commonly used in software engineering to describe additional costs of rework after implementing a quick solution) with most of the machine learning applications when manual workflows are applied. A concept to mitigate these costs is MLOps which aims to automate steps from development to production and enable an efficient governance over the machine learning workflow (Ruf, Madan, Reich, Ould-Abdeslam 2021). Ruf, Madan, Reich and Ould-Abdeslam lay out the workflow stages of MLOps in the following four phases: Data Management, ML Preparation, ML training and Deployment phase (2021) and define various requirements for tools that support MLOps as seen in the following image.

Figure 3: Requirements for tools supporting MLOps (own illustration summarizing the work of Ruf, Madan, Reich, Ould-Abdeslam 2021)

They further name general requirements such as scalability or user friendliness (Ruf, Madan, Reich, Ould-Abdeslam 2021, p. 22 f). Benefits of MLOps are particularly realized whenever organizations manage multiple models and datasets (Mäkinen, Skogström, Laaksonen, Mikkonen 2021). Most companies still are finding out how to best use their data or build the first PoCs but will progress to have multiple models and versions of them in the future (Mäkinen, Skogström, Laaksonen, Mikkonen 2021), which is why MLOps can be seen as a developing topic with an interesting future both in academia as in practice. Summarizing, MLOps is a field of research and practices that greatly influences the adoption and deployment of machine learning techniques by making the process more efficient and controllable. AutoML can help in multiple ways along the MLOps phases depicted above, we will explore how in the following chapters.

3.3. Automated Machine Learning (AutoML)

Hutter, Kotthoff and Vanschoren (2018) classify AutoML as the democratization of Machine Learning because it makes the powerful technology available to users from every background. The promise of AutoML concerning enterprise adoption is that it could empower domain experts in building their own ML models and therefore free resources of data scientists for them to focus on the projects that look most promising and drive business value (Carlsson et al. 2020, p. 2). AutoML does that by automating each step along the development of machine learning models, from data preparation, feature engineering, algorithm selection and hyperparameter tuning to model validation (Carlsson et al. 2020, p. 4).

AutoML is not only used in every step of the Machine Learning workflow, but it also spans all categories of machine learning with tabular data, from (multi-label) classification (Wever, Tornede, Mohr, Hullermeier 2021), clustering (Poulakis, Doulkeridis, Kyriazis 2020) to even tuning deep neural networks in a discipline called Neural Architecture Search (NAS, see: Elsken, Metzen and Hutter 2019). AutoML systems regularly perform better than human data scientists in these tasks (Purwanto, Pal, Blair and Jha 2021). The principles of AutoML are furthermore applied to non-tabular data, leading to AutoML systems performing complex tasks such as text classification (Brandle, Hanussek, Blohm and Kintz 2021). Even more astonishingly, these systems already are able to outperform humans in such complex tasks as well (Blohm, Hanussek and Kintz 2021).

3.3.1. Explainable (Auto)ML and humans in the loop

A Machine Learning system can only provide business value if humans trust, understand and learn to interpret its outcome (Kennet and Redman 2019). This is especially true for AutoML systems, where domain experts interact with the system instead of technologists (Karmaker et al. 2018). Studies have found that a higher degree of explainability in AI systems increases trust in outcomes and makes accessing the information and interpreting it easier for users (Shin 2021). There are ways to ensure explainability, for example by using only certain algorithms: Belle and Papantonis (2021) compared and rated Machine Learning algorithms according to their level of explainability and called i.e. Decision Trees transparent models as their structure allows a user to clearly understand the decision-rules of the algorithm. With this knowledge it would be possible to allow users of an AutoML system to preselect the level of explainability that is necessary for a problem at hand and trough that

On the other hand, at the current level of automation, AutoML systems can only unfold their full potential if humans are still interacting with the systems as we haven't reached a fully automated ML process yet (Karmaker et al. 2018). This brings the so-called human in the loop narrative to the table, which lives of a “tension between speed and human oversight” (Crisand and Fiore-Gartland 2021, p.1). The basic question is to which level a human must and should be involved in an AutoML workflow, weighing explainability and performance. Some tasks just like a proper problem formulation for the data science project at hand are impossible to properly automate considering the current maturity of the AutoML technology (Karmaker et al. 2018). More strategic tasks such as applying ML in the right context where it fosters the most business value and taking coherent action from interpreting ML outcomes, will need a “human in the loop” in the foreseeable future.

3.3.2. AutoML and Enterprises

With AutoML, Gualtieri etl a. (2020) estimate that data science teams in enterprises will be able to implement eight times the number of use cases as before (p. 2), showing the massive value that lies in AutoML adoption. This is especially valuable as it mitigates core machine learning adoption challenges (Lee and Shin 2020) by making the scarce resource of data science experts focus on the cases that drive business value and select these valuable cases trough building a big number of proof of concepts (PoCs) with AutoML. But what kinds of tools are out there to do that? Carlsson et al. (2020) differentiate between four different types of AutoML solutions, namely:

- Automation-focused machine learning platforms, trying to minimize the need for human intervention as much as possible (e.g. DataRobot)
- Multimodal machine learning platforms, offering a wide range of functionalities and keep humans in the loop (e.g. Dataiku)
- Deep learning focused AutoML solutions, which work with humans especially to make complex tasks as image or text data work (e.g. Google AutoML) and
- Augmented business intelligence solutions, which give automated insights and lead humans to follow-up on interesting or unusual patterns (e.g. TIBCO) (p. 5).

In coherence with the research question and objectives it is not the goal of this thesis to reinvent AutoML tools but to specifically analyse how a tool must be designed to include non-technical individuals to collaborate on and leverage AutoML techniques in their daily lives. There seem to be enough tools out there already that cover AutoML tasks and the proposal in this thesis will rather be a workflow that builds on these existing tools and technologies to foster AutoML usage by non-technical users. Looking at the internet representations of tools like Dataiku[1] and DataRobot[2] we can see they extensively support the whole process of applying machine learning techniques like data preparation, feature discovery, applying AutoML for tabular, text and image data, MLOps and continuous AI/ Data operations techniques. Further analysing these tools it seems that they are not built with the same non-technical user in mind as it will be in this thesis: steps like preparing the data are indeed provided in a way that users do not know how to code, but they need to know how they want their data to be prepared for the model they want to build. Do non-technical users know that

[1] https://www.dataiku.com/product/key-capabilities/
[2] https://www.datarobot.com/platform/

Both last accessed: 17.10.2022, 4 PM

or is it better to include experts into the workflow here? Is it the best way to ask them to do that in the beginning of the process or develop the data model with different iterations? This is where this thesis will have its contribution. It asks: what is the sequence of steps that allows an organization and its non-technical users to best leverage AutoML?

4. ENVIRONMENT ANALYSIS – MANAGEMENT CONSULTING INDUSTRY

The Management Consulting industry has been chosen as an optimal example to study the workings of AutoML in enterprises on account of its reach into all industries and wide array of tasks therein. Diving deep into the science around Management Consulting (macro-perspective) allowed me to reflect on my practical experience as I work in this profession myself and can observe behaviour on a micro-level. To allow you as a reader to experience both perspectives (micro and macro), I will try to include illustrative examples of observed individual actions when extracting the big picture from literature.

4.1. THE MANAGEMENT CONSULTING INDUSTRY

4.1.1. Definition and Purpose

Just defining Management Consulting (MC) is a hard task in itself as different authors see different scopes and deliminations to other (consulting) practices. Although the industry already developed in the 19th century (McKenna 1995), the profession still has "ambiguous boundaries" (Jerónimo, Pereira and Sousa 2019) and crosses many other industries and subjects. The following table compares common definitions of MC from different sources in academic literature selected out of extensive research in the biggest scientific databases with definitions of practitioners and representing organizations.

Source	Definition of Management Consulting/ Management Consultants
Jerónimo, Pereira and Sousa (2019, p. 1)	"[...]include any activity that has as its apparent justification the provision of some kind of support in identifying or dealing with management problems, provided by individuals, groups, or organizations that are external to the particular management domain and which are contracted by the management on a temporary basis."
Greiner and Metzger (1983)	"Management consulting is an advisory service contracted for and provided to organizations by specially trained and qualified persons who assist, in an objective and independent manner, the client organization to identify management problems, analyse such problems, and help, when requested, in the implementation of solutions"
Canback (1998), p. 3	"[...] those who provide general management advice within a strategic, organizational or operational context, and who are institutionally organized in firms."
European Federation of Management Consultancies Associations: FEACO (2021)	"Management consulting covers a wide range of services and can be defined as independent advice and support on management issues. [...] The management consultant is therefore a provider of help, a supporter, a spokesperson, but also a seizer of opportunities, a problem solver and a decision maker."

Coming from the variety of definitions it is helpful to take a closer look at the services provided by MC companies to obtain a better understanding of the industry and its purpose. A comprehensive segmentation of MC services is for example given by Curusku (2018), who differentiates between Business Strategy, Marketing Management, Operations and Value Chain Management, Financial Management, Human Resource Management and other services (which include projects that span across multiple categories as well as rather specific topics like accounting, governmental programs, or technology development), while also stating that the "distinction between management consulting services and more technical IT consulting services is becoming less and less evident" (p. 2). From a more general perspective on what management consultants do, Turner (1982) listed eight different tasks of MC in a hierarchical order:

1. Providing information to a client
2. Solving a client's problem
3. Making a diagnosis, which may necessitate redefinition of the problem
4. Making recommendations based on the diagnosis
5. Assisting with implementation of recommended actions
6. Building a consensus and commitment around corrective action
7. Facilitating client learning
8. Permanently improving organizational effectiveness.

While this list from around 40 years ago still proofs surprisingly consistent with more current research and practice, the hierarchy and the way these tasks are delivered today have undergone changes since MC companies adapt their own business models accordingly to the digital disruption their clients face trough specialization and investments in technological expertise (Jerónimo, Pereira and Sousa 2019).

4.1.2. Market overview and development

According to IBISWorld there are close to 2,4 million MC firms globally employing over 5,5 million people (Global Management Consultants Industry - Market Research Report, 2021). In Europe alone there are over 880.000 MC companies and close to 1,5 million people employed in the industry (Management Consultants in Europe - Market Research Report, 2021). Although being a fragmented market with low concentration on few major firms (Curusku 2018, p. 13), some incumbents stick out when looking at the MC industry. There are on the one hand three big strategy-focused consulting firms McKinsey, Boston Consulting Group (BCG) and Bain, often abbreviated with MBB. On the other hand, there are the four huge accounting firms that grew their consulting business more and more over the last years, namely Deloitte, KPMG, PwC and EY, summarized with "The Big 4". These major players have been joined by more IT-driven firms such as IBM or Accenture. Besides these huge companies, there is a variety of specialized consulting firms for specific industries or services and increasingly startups and independent freelancers join the MC market (Curusku 2018, p. 15). In the US as much as 45 % of new management consultants are freelancers or startups (Edwards 2016). Other estimations for European countries range between 13 % and 31 % freelancers of overall consultants with an increasing percentage as more and more newcomers work independently (Hardt 2018, p. 389).

Within the last decade, consulting related to digital transformation has come to be the fastest growing type of service for consulting companies in Europe (Tavoletti, Kazermargi, Cerrutti, Grieco and Appolloni 2021, p. 2). And more recently, a specific part of digital transformation gained traction as "Consultancies Are Answering Clients' Call To Realize The Value Of AI" (Granzen et al. 2019, p. 2). It can be estimated that nearly half of global AI projects are done with external service providers such as MCCs (Goetz, Leganza, Granzen, Hennig 2021 a). When looking at the strongest AI consultancies according to Forrester research, we find the typical Management Consultancies having a leading market position as PwC, KPMG, McKinsey and BCG make up the leaderboard (Goetz, Leganza, Granzen, Hennig 2021 b). In the following chapter we will also explore how these changes reflect in MCCs business models.

4.1.3. Challenges and opportunities: MC about to be disrupted?

Succeeding, I want to discuss strengths, risks, threats, and opportunities of MCCs and end this chapter with a high-level, general SWOT analysis for the MC market. That allows me to present a clear analytical result which sets up the definition of requirements later in this thesis. The drastic question of this part could also sound like: Are MCCs about to fall victim to creative destruction (Schumpeter 1950) in a globalized competition? Or, and if yes how, will they strive from the AI revolution (Harari 2017)?

Wang and van Bever stated in a widely noticed article in Harvard Business Review (2013) that consulting is on the cusp of disruption. They argue that the competitive advantages of MC firms disappear in an increasingly digital and more transparent world, which forces large incumbents to rethink and extend their business model (Christensen, Wang and van Bever 2013). The authors name four major implications for MC firms, one of them being the increasing importance of data analytics technology (Christensen, Wang and van Bever 2013, p. 9). Curusku (2018) also highlights the use of (big) data analytics capacities as one of the biggest opportunities and threats for today`s MC firms, while Tavoletti et al. (2021) point out that MC companies already try to build such capabilities through hiring strategies and mergers and acquisitions. Other authors blow into the same horn stating that more than 85 % of interviewed experts (n=15) see technological changes as the future of consulting (Jerónimo, Pereira and Sousa 2019). A widely noted example for how the future could look like for consulting is given by MCKinsey solutions, an asset-based (e.g. subscription models) consulting business model that offers data-driven insights from industry-specific tools (compare Christensen, Wang and van Bever 2013 and Curusku 2018).

Generally, consultancies are adopting their business models and identity in many ways to clients' needs and the technological advancements. They integrate digital assets into every project and use them to drive the value for their clients, recently especially in AI and analytics (Tavoletti, Kazermargi, Cerrutti, Grieco and Appolloni 2021, p. 9ff). They furthermore offer end-to-end solutions (e.g. from finding a use case to deploying an assisting AI application) that are increasingly often delivered in a network of experts, in which consultants play a coordinating role (Tavoletti, Kazermargi, Cerrutti, Grieco and Appolloni 2021, p. 9 ff). Underscoring the observation of MCCs opening up to networks and ecosystems in AI consulting, researchers have found increasing evidence of cooperation between MC incumbents and innovative or very technical service providers and technology platforms (Flynn and Kowalkiewicz 2018, p. 109). The usual starting point for consultancies when engaging in an end-to-end AI project with a client is to start with a proof of concept (PoC) or a pilot project to make added value visible to clients (Granzen et al. 2019, p. 3). An opportunity arises when these PoCs move to production and allow for subscription-based, profit-sharing or other outcome-based business models (Tavoletti, Kazermargi, Cerrutti, Grieco and Appolloni 2021. P. 11). Flynn and Kowalkiewicz furthermore point out that the level of scalability of analytics and algorithmic solutions dramatically exceed the level of scalability of humans – posing both a threat and an opportunity to MCCs (2018, p. 102). MCCs already show willingness to implement digital technologies in order to augment human skills and activities and allow for new business models that prioritize scalability (Flynn and Kowalkiewicz 2018, p. 108). Werth and Greff (2018) created a framework for digital business models in consulting which clarifies the points mentioned so far.

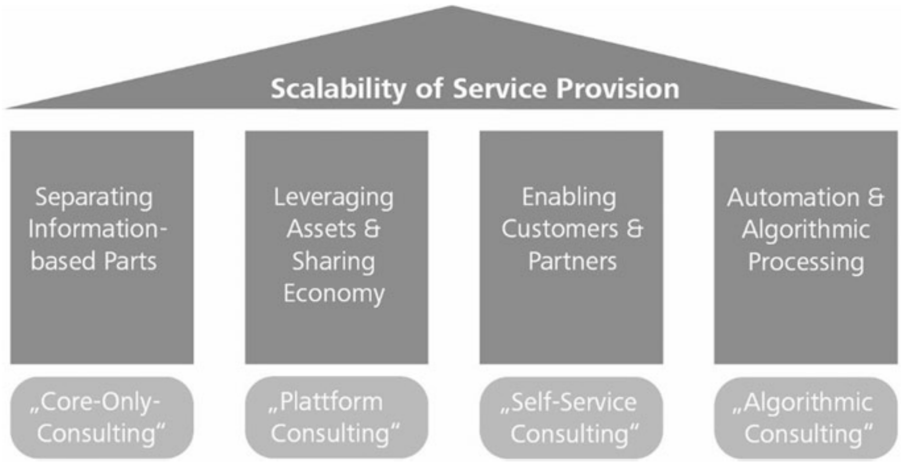


Figure 4: Scalability for consulting services by Werth and Greff (2018, p. 125)

Behind every successful AI project stands a clearly defined business need and the application of the right technology in the right context (Kennet and Redman 2019). Curusku (2018) expresses something similar when he says: “Even with big data, traditional business consultants are needed to ask the right questions” (p. 19). Compared to start-ups and tech-companies that try to augment or virtualize (compare Nissen 2019) the need for consulting purely with technology, traditional consultancies have a competitive advantage as they historically come from the business/ strategy domain and are now marrying this expertise with AI approaches to enable multidisciplinary service models (Granzen et al. 2019). These service models even have the potential to allow for altered interactions with new customers (Nissen 2018, p. 16).

Especially in an environment of multidisciplinarity, access to highly skilled workforce is a success factor in consulting (Curusku 2018, p. 11). This is easily comprehendible as people really are the product of consulting companies in a classical business model. But looking at the fact that data scientists are one of the scarcest resources on the labour market (Jarvis 2020), that poses a threat to MCCs as clients increasingly expect data driven advisory (Nissen 2018, p. 3). Transformative business models like asset-based consulting might be less dependent on selling well educated people per hour. But still companies need the right (technical) talent to build such assets. MCCs want to bridge this gap through mergers and acquisitions (Tavoletti, Kazermargi, Cerrutti, Grieco and Appolloni 2021). In cases when MCCs face a lack of resources or very specific knowledge they tend to reach out to freelancers or independent niche consultants (Hardt 2018, p. 391). That further points out the opportunity of incumbent MCCs to play a coordinating role in a network of experts. An advantage of classical MCCs compared to newcomers and start-ups on the consulting market is to have a head start in gaining clients' trust which is of highest importance when handling for example sensible data for an AI application (Flynn and Kowalkiewicz 2018). But in the war for talent, MCCs must be highly adaptive and include newcomers and start-ups to preserve a dominant role in the consulting ecosystem.

The following SWOT analysis summarizes the strengths, weaknesses, opportunities, and threats of MCCs in a world of digital disruption.

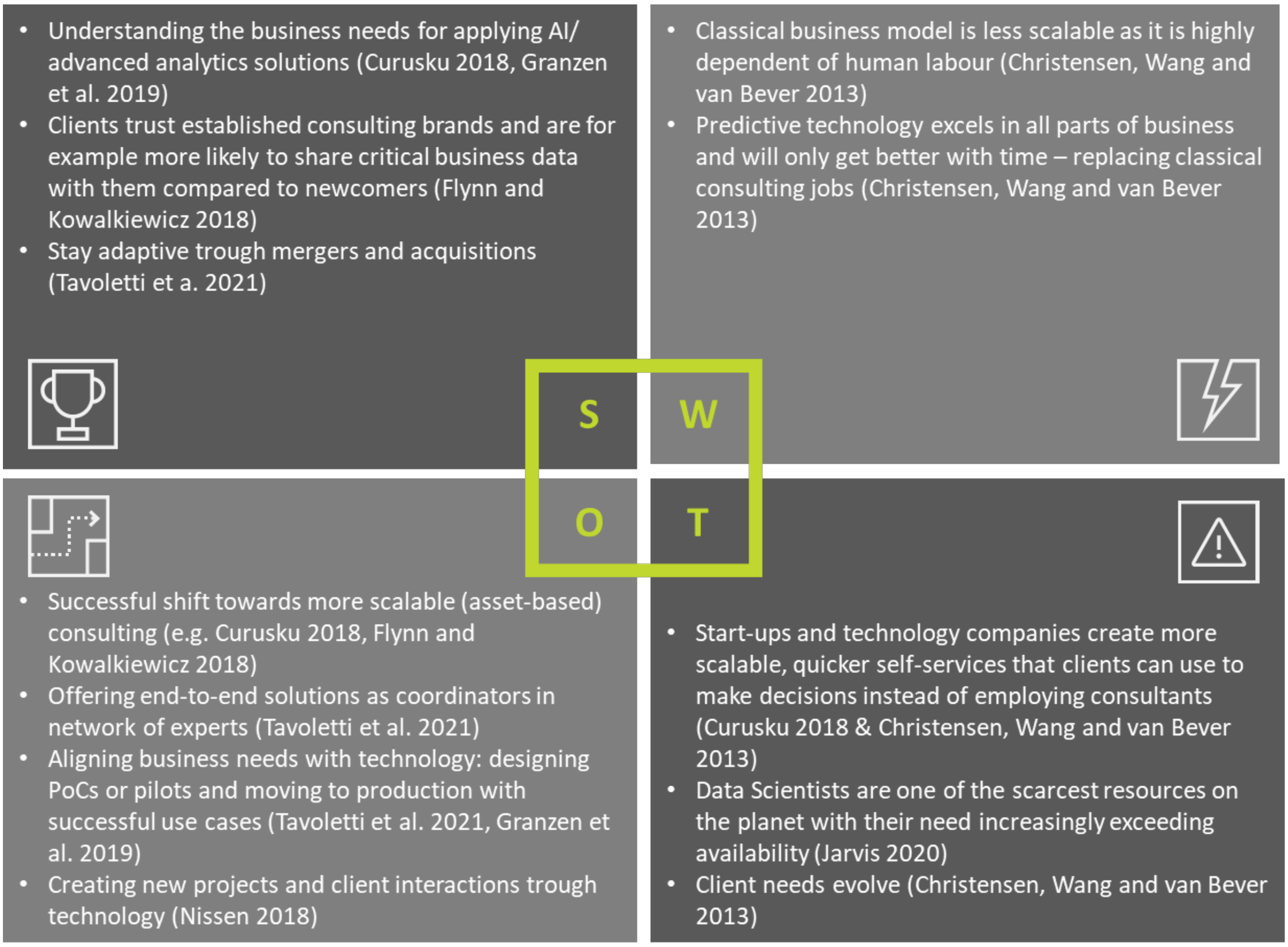


Figure 4: SWOT-Analysis Management Consulting industry (own illustration)

5. PROPOSAL

5.1. THEORY: SYNTHESIS OF THE ENVIRONMENT AND RIGOR CYCLE

Aggregating the learnings from the literature review on (automated) machine learning and the problem environment of Management Consulting Companies a theoretical framework for abstracting further requirements for a possible solution is laid in this chapter.

The first core learning is that AI and machine learning have become mature technologies with increasing applications in all industries and sectors (Borges, Laurindo, Spínola, Goncalves and Mattos 2021). Still, most companies seem to be in the beginning of their journey concerning machine learning and AI adoption into all processes and services: Accenture guesses that 80 to 85 % of companies are stuck in the proof of concept phase of their AI efforts (Reilly, Depa and Douglas 2019). The typical process to implement machine learning solutions involves starting a proof of concept and/ or pilot project, evaluate the ratio of possible business value versus costs and according to that move to production or not. Business value only realizes if there is trust in results of a ML system, which highlights the importance of security and interpretability of models and often is ensured by keeping human experts in the loop (Kennet and Redman 2019).

To understand what hinders and fosters the implementation of further ML systems we can learn from MCCs which are (together with other external service providers) involved in around half of all AI projects executed in organizations (Goetz, Leganza, Granzen, Hennig 2021 a). MCCs usually act as coordinators in a network of service providers that enable an end to end (i.e., from finding a use case to realizing value) experience for a client organization (Flynn and Kowalkiewicz 2018). Core value propositions of MCCs include defining use cases and assessing business value while it is hard for them (as for most other organizations) to bind talent in the field of data science as there is an overall shortness (Curusku 2018). This observation is even more important considering that the MCC market heads towards being more and more fragmented and complex with increasing numbers of freelancers and more technical companies joining the market (Edwards 2016 and Hardt 2018) as well as classical MCCs adopting their business model towards asset-based consulting (Curusku 2018, Christensen; Wang and van Bever 2013; Jerónimo, Pereira and Sousa 2019).

AutoML can mitigate some of the challenges MCCs face when adapting their business model towards asset-based consulting with ML systems. With AutoML MCCs can deliver pilots or proof of concepts at a significantly faster rate by allowing everybody to use the technology and speeding up the process of developing first models (Gualtieri et al. 2020). Further, AutoML can be used by ML experts to finetune algorithms and neural networks (Hutter, Kotthoff and Vanschoren 2018). Other steps of the process to implement ML systems still seem to rely on human expertise: MCCs can help define use cases for client organizations as well as evaluate them concerning business value by drawing from their industry expertise (Tavoletti, Kazermargi, Cerrutti, Grieco and Appolloni 2021). Once ML systems move from pilot stage to production, MCCs can realize value using MLOps techniques to automate workflows in development and integration of models (Ruf, Madan, Reich, Ould-Abdeslam 2021) which enables them to transform towards subscription-based business models.

MCCs are put into a position where they coordinate the different roles within an ecosystem according to their strengths and value proposition (Tavoletti, Kazermargi, Cerrutti, Grieco and Appolloni 2021, p. 9 ff).

The following illustration shows the role of MCCs in the process of using machine learning techniques and highlights where AutoML is applied.

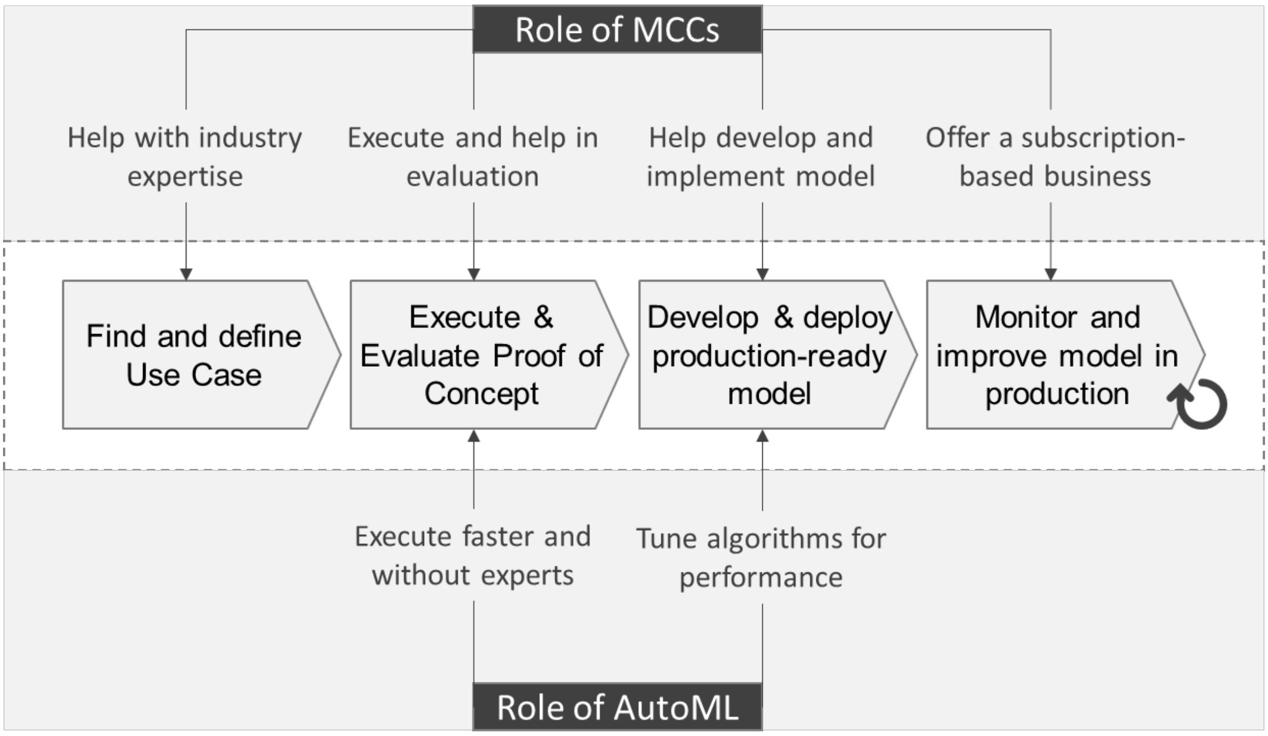


Figure 5: Role of MCCs and AutoML in applying machine learning techniques (own illustration)

Summarizing on a high level, an information system must fulfill the following assumptions to empower MCCs and their non-technical individuals to collaborate on and leverage AutoML techniques in their daily lives to unlock the value of available data for their clients:

- The proposal must allow its users to define relevant use cases for using machine learning in a business context and launch fast Proof of Concepts (PoCs).
- The proposal must enable its users to execute and explain PoCs in order to evaluate the possible business value of a model.
- The proposal must support collaboration between business users, machine learning experts and different organizations to deploy and monitor the best use cases for business value realization.

These high-level assumptions mirror the most important steps of the end-to-end process in deploying a machine learning model and serve as our theoretical guiding principles to derive requirements for a proposal.

Coming from the high-level requirements it is now possible to describe a scenario of how individuals within a MCC might use an AutoML tool in their daily lives to create value for their clients (Use Case):

A group of Management Consultants without knowledge in Data Science might face a strategic issue with a client in any industry. An example for such an issue would be trying to better address different target groups within the client's customer base to cross- or upsell. Helping to solve this issue, the team of Management Consultants wants to apply a Machine Learning algorithm to delimit different Personas within the customer base from each other. They use a tool (this proposal) to conduct a Proof of Concept where the client's data is preprocessed and analyzed without needing to code by using AutoML. The team further uses the created Personas in a Design Thinking workshop where they ideate and test different ideas to best address each cluster of customers on a targeted basis. The client likes the approach and sees the value created from the Machine Learning technique. Now the Management Consulting team has an opportunity to upsell. They offer their client a service on a subscription basis where they update the cluster analysis monthly and give recommendations on how to target individual customer clusters. The client subscribes. The team of Management Consultants now need technical expertise to implement the Machine Learning algorithm in such a way that it requires minimal manual effort to update the analysis to maximize profit. They share the PoC with technical experts from within or outside their company. These experts also finetune the algorithm to maximize its performance and continuously improve it using MLOps techniques. Other companies from within the client's industry might also subscribe to the service initially created for the root client. The Management Consulting Company now possibly created large profits by using a tool and without having to hire additional technical talent as the ones they have can focus on value-proven cases exclusively.

The following chapter will detail the high-level and the described Use Case into lower-level assumptions which serve as basic requirements for the proposal to be built.

5.2. Requirements for a possible solution

Coming from the theory laid out in the chapter before and by utilizing the description of a possible Use Case, more detailed requirements for the proposal can be derived. The requirements will be presented along the three high-level assumptions and are decomposed from the described Use Case. A Best Practice from the field of software engineering is to delimit functional from non-functional requirements (Pressman 2009, compare chapter 2.2).

The following table shows the functional requirements (F1.1 – F3.2) derived from high-level assumptions:

REQUIREMENT ID	REQUIREMENTS
HIGH-LEVEL ASSUMPTION 1: THE PROPOSAL MUST ALLOW ITS USERS TO DEFINE RELEVANT USE CASES FOR USING MACHINE LEARNING IN A BUSINESS CONTEXT AND LAUNCH FAST PROOF OF CONCEPTS.	
F1.1	The system must help users identify possible use cases.
F1.2	The system must help users to identify the data needs of a use case to better define it.
F1.3	The system must be able to import training data from different sources.
F1.4	The system must help users to prepare training data to make them readable and useful for machine learning algorithms.
F.1.5	The system must be able to execute different AutoML technologies.
HIGH-LEVEL ASSUMPTION 2: THE PROPOSAL MUST ENABLE ITS USERS TO EXECUTE AND EXPLAIN POCS IN ORDER TO EVALUATE THE POSSIBLE BUSINESS VALUE OF A MODEL.	
F2.1	The system must enable non-technical users to explain the technology behind the chosen use case, so it fosters trust in the outcomes.
F2.2	The system must allow users to adopt parameters in the execution of a use case according to their specific business needs.
F2.3	The system must allow the user to explain how data is kept and security of training data is provided to foster trust.
F2.4	The system must enable the user to explain the outputs of the AutoML system in a way that they can be translated to business value.
HIGH-LEVEL REQUIREMENT 3: THE PROPOSAL MUST SUPPORT COLLABORATION BETWEEN BUSINESS USERS, MACHINE LEARNING EXPERTS AND DIFFERENT ORGANIZATIONS TO DEPLOY AND MONITOR THE BEST USE CASES FOR BUSINESS VALUE REALIZATION	
F3.1	The system must help machine learning experts to build production-ready models faster after a successful proof of concept.
F3.2	The system must allow MCCs and their client organizations to perform MLOps tasks (eg model versioning, deployment, and monitoring)

Table 1: Functional requirements (F1.1 – F3.2) derived from high-level assumptions

Next to detailing the assumptions into functional requirements it is important to understand overarching design principles for a possible AutoML tool that can help non-technical individuals integrate machine learning in their daily lives. These are non-functional requirements for the proposal being developed (NF1.1-NF1.3).

As it was established that typical users of the tool would not have a background in computer science or other technical fields,

> NF1.1 the proposal should never display information that is too technical or requires explicit knowledge on the process and methods of data science from a business-user.

This non-functional requirement leads us to various implications for translating the functional requirements into a proposal. For example, one functional requirement is to help users define an applicable ML Use Case (F1.1). There are certain ways to approach that, one of them being to provide a database with possible Use Cases that a user can search and be inspired by. These Use Cases must be presented in an easily understandable and non-technical wording that allows business users to understand how to create value from a use case. Another functional requirement is that the proposal must be able to execute different AutoML technologies (F1.5). The proposal might fulfill this requirement by integrating existing AutoML technologies (for example open-source technology or chargeable APIs to existing technology providers) but NF 1.1 tells us that the proposal must be designed in a way that no information is displayed that is not necessary or comprehendible to the business user.

Business users can easily forward successful proof of concepts to experts within the tool so that they can prepare the PoC for production (F3.1). Experts receive detailed information about the AutoML model which are not displayed to business users but help experts to perform their tasks faster (for example information on feature importance, algorithm performance etc). This still complies with NF1.1 as the intended user has a more technical education in this step.

But even the easiest language is not enough if non-technical users do not know which steps, they need to take to achieve a certain result. That leads to non-functional requirement 2:

> NF1.2 The proposal must guide users through an end-to-end workflow that begins with defining the use case to evaluating the first model.

An example how this non-functional requirement influences the proposal: Thinking about the Use Case scenario described in the chapter before the Management Consulting team's (business users) journey starts with receiving data from their client. Some basic steps of data preparation and feature engineering now must be applied assisted by AutoML to get the most out of the machine learning techniques. This step must be initiated automatically as soon as input data is provided so inexperienced business users cannot skip this important part of developing machine learning models. Another example of how NF1.2 influences the proposal might occur after the AutoML analysis is conducted: This is when the MCC can convince its client of the value of the machine learning model and sell a subscription-based service. Since MCCs usually present their information in Microsoft Power Point slides or similar presentation tools, the proposal automatically prepares slide templates that help to explain outputs of the machine learning model and their value. Further, after finishing a PoC in the proposal, documentation on data privacy and security can automatically be provided to the user.

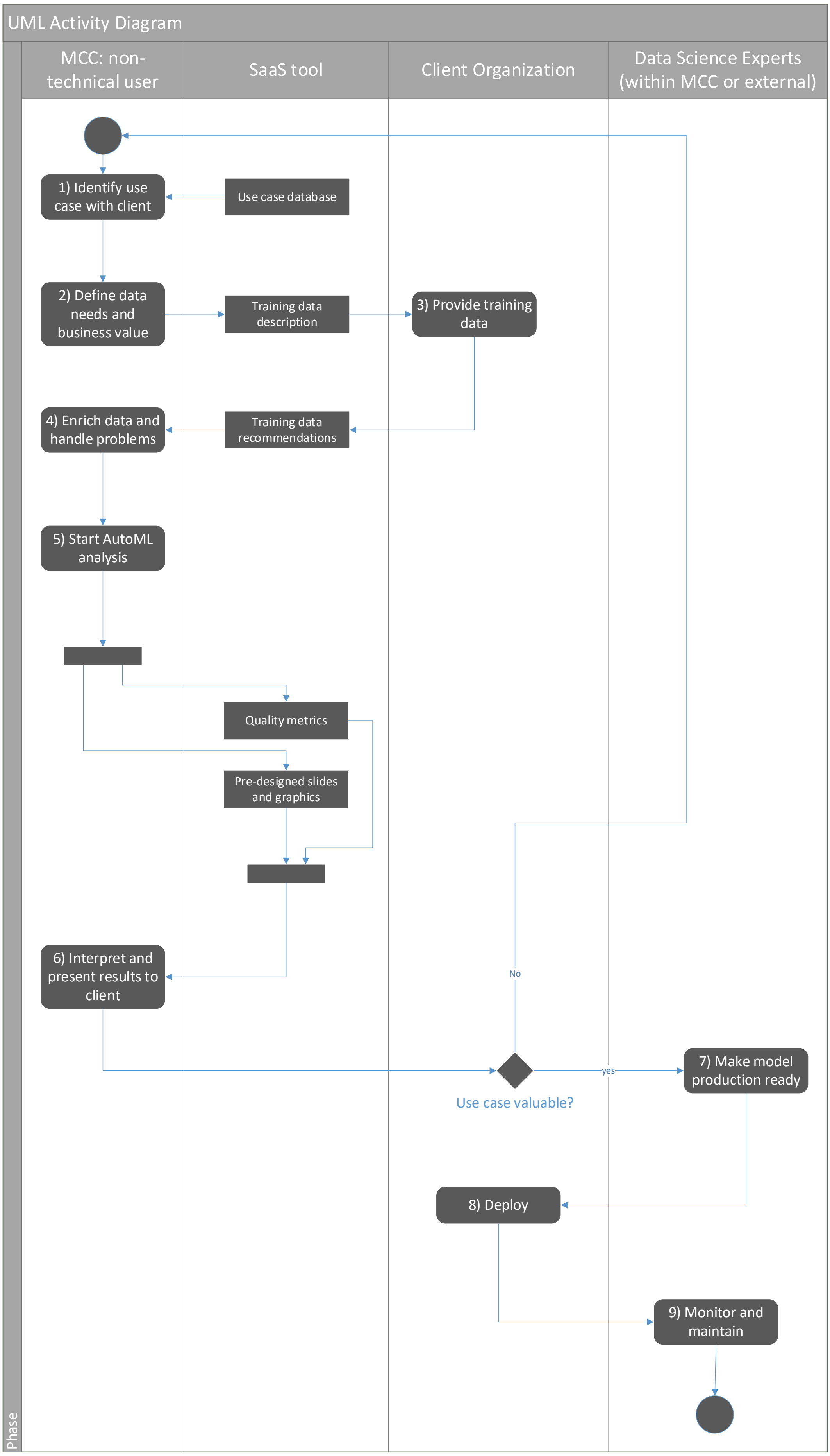
We learned that trust in the outputs of a machine learning model is of paramount importance for their ability to be translated into decisions and actions. Therefore, it is important that

NF 1.3 the proposal must display relevant explanations to the user at every step of the workflow so the user understands the process well enough to trust its results.

This non-functional requirement reflects in the proposal for example after selecting a use case in the form of explanations of the underlying machine learning technique that are used (e.g. classification, regression, clustering).

5.3. PROPOSAL FOR A WORKFLOW IN AN AUTOML-POWERED INFORMATION SYSTEM

The proposal consists of a workflow represented through a UML activity diagram. A.UML activity diagram “depicts the dynamic behaviour of a system (…) through the flow of control between actions that the system performs” (Pressman 2009 p.853). The reason for choosing a UML activity diagram as my proposal is laid out in chapter 2.2. A UML activity diagram can be used in combination with swimlanes that indicate who or what executes a certain step of the workflow (Pressman2009 p.853). According to our use case description we have MCCs and their non-technical users as one actor, the client organization as a second actor and data science experts that might work within the MCC or come from another organization as third actor. I decided to also include the system itself as an actor. Rectangles with rounded corners represent actions that are executed by the actor in which lane they are sorted (Pressman 2009). One may also include rectangles with sharp edges into a UML activity diagram. They represent objects or data that flow trough the workflow (Sommerville 2010 p. 123). Arrows between actions or objects indicate the direction of the flow. Forks and joins are represented trough horizontal bars and indicate the separation of activities if they are carried out in parallel and their rejoining when entering a sequential flow again (Pressman 2009). Decision nodes represent different options for the flow of control to continue. The nodes indicate the beginning and the end of the workflow. I numbered the workflow-steps so that they can be easily referred to in the proceeding chapters. They will be referred to as WF1 – WF9.


UML Activity Diagram
MCC: non-technical user
SaaS tool
Client Organization
Data Science Experts (within MCC or external)
1) Identify use case with client
Use case database
2) Define data needs and business value
Training data description
3) Provide training data
4) Enrich data and handle problems
Training data recommendations
5) Start AutoML analysis
Quality metrics
Pre-designed slides and graphics
6) Interpret and present results to client
No
yes
Use case valuable?
7) Make model production ready
8) Deploy
9) Monitor and maintain
Phase

5.4. DEMONSTRATION

As an instantiation of the proposal given in the chapter before, user interfaces were built to demonstrate how the workflow would interact with a possible user. The demonstration consists of 5 interfaces that emulate the user's experience while following the workflow (proposal). The interfaces were designed in Microsoft Power Point under consideration of the functional and non-functional requirements derived from theory (chapter 5.2) and following the steps of the proposed workflow (chapter 5.3). As the goal is to get high-quality feedback for the proposal it is important that the demonstration allows experts to imagine the proposed workflow being applied in a real-world context. Thus, the demonstration is set up to look like an existing Software-as-a-Service (SaaS) tool that leads users through the proposed workflow and lets them interact with it. The demonstration therefore features a made-up name, "pocket.ml", which should spark associations with putting machine learning techniques into one's "pocket" and integrating it in its user's daily lives. Following, I will describe each interface briefly and then give an overview which requirements and steps of the proposed Workflow are depicted for demonstration purposes.

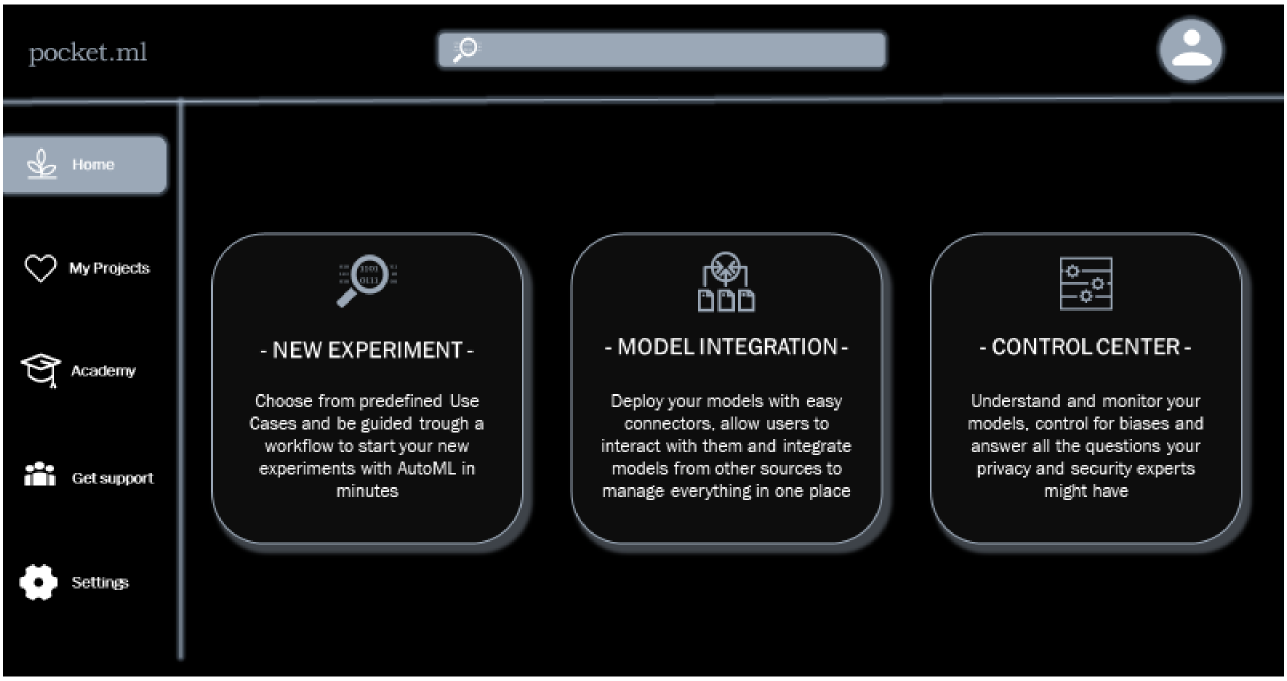


Figure 6: Demonstration page 1

On the first page, the user can choose to conduct a new experiment (i.e. to create a PoC for a machine learning model with AutoML). Further, there are options for "Model Integration" and a "Control Center". The "Model Integration" option is for value-proven PoCs that a user wants to move to production. Typically, an expert will conduct this task after reviewing and finetuning the model created by AutoML. A possible tool might help with providing standard connectors to legacy systems and give different deployment options in the cloud (e.g. docker images). The "Control Center" option helps to overview model performance and answers privacy and security questions. Both options therefore help with executing MLOps tasks. The demonstration focuses on conducting a new experiment while the options "Model Integration" and "Control Center" will not be explored any further in this demonstration and serve the purpose of giving reviewing experts a feeling for what could be the range of functionalities of an actual SaaS tool that is built after the proposal. On the left side, the first page of the proposal also lists some typical menu items of any SaaS tool to make the demonstration feel more real: the home-screen is selected right now, the "My Projects" tab might give an overview over past experiments, the "Academy" provides resources to learn handling the tool better like videos and documentations, the "Get Support" tab lets you contact customer support and your preferences are manged in the "Settings" tab. On the top of the page there is the name of the fictious SaaS-tool, a search bar and a user icon which on clicking on it might provide the option to log out or change users.

While we will explore the further workflow after clicking on "New Experiment" on the first page, the other options especially help to fulfil the following requirements and steps of the workflow.

REFERENCE	REQUIREMENT/ WORKFLOW-STEP
F3.1	The system must help machine learning experts to build production-ready models faster after a successful proof of concept.
F3.2	The system must allow MCCs and their client organizations to perform MLOps tasks (eg model versioning, deployment, and monitoring)
WF8	Deploy
WF9	Monitor and Maintain

Table 2: Requirements and Workflow-steps integrated in page 1 of the demonstration

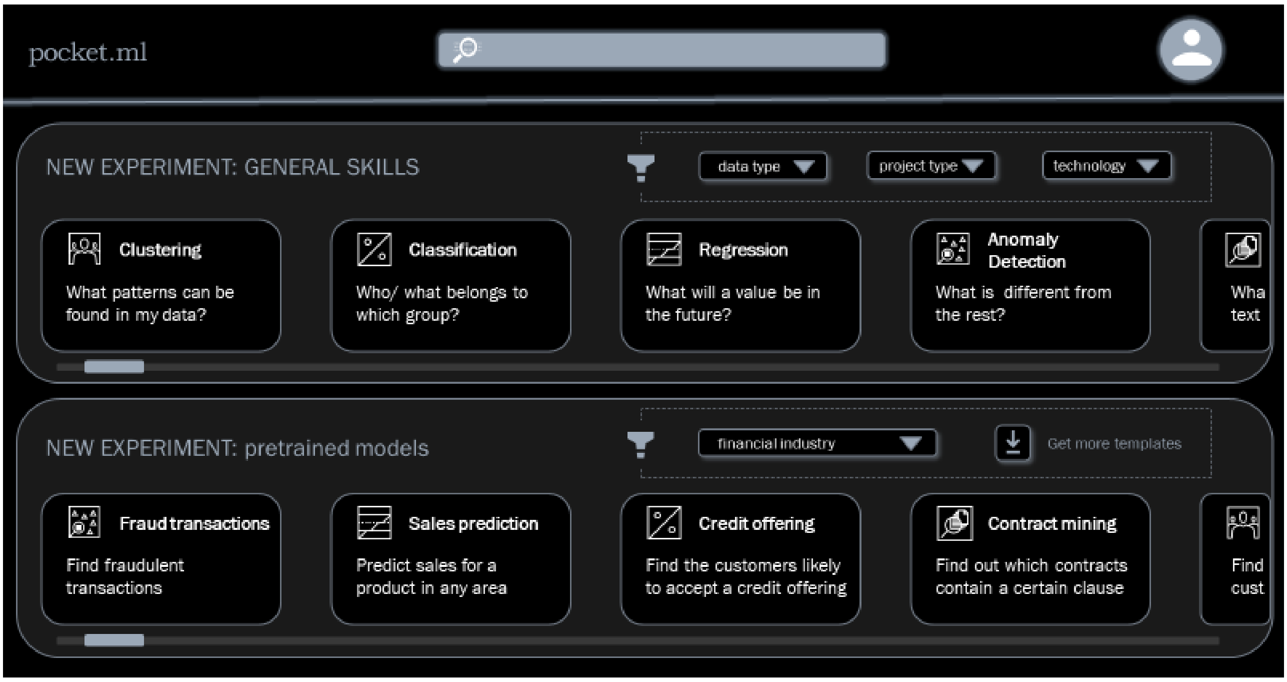


Figure 7: Demonstration page 2

On the second page of the demonstration the user is assisted in defining the use case for the new experiment. In the upper paanel, general use cases are listed by technology (e.g. classification, regression). To comply with NF1.1 there are questions formulated in non-technical language to assist non-technical users in finding the right use case. In the bottom panel use cases are specified per industry which can be selected in the upper right part of the panel. Industry-specific use cases might make it easier for users to identify what is relevant for their needs. The tempates for these use cases might come from past projects shared within a company or even across companies (MCCs).

REFERENCE	**REQUIREMENT/ WORKFLOW-STEP**
F1.1	The system must help users identify possible use cases
NF1.1	the proposal should never display information that is too technical or requires explicit knowledge on the process and methods of data science from its user
WF1	Identify use case with client

Table 3: Requirements and Workflow-steps integrated in page 2 of the demonstration

Figure 8: Demonstration page 3

In the third page of the demonstration we already know which use case the user wants to execute. For the purpose of this demonstration the use case that will be executed is clustering customer data. As it was established that trust in an analysis is important for the actionability of the results and trust comes from understanding the analysis in depth, the technology (here: clusteirng) is explained on the left side of the interace. Also, a description of the use case can be dowloaded on the page and customized to the specific appliation context. Following the porposed workflow it is now important to qualify a non-technical user to describe and select the needed data to execute the use case (in this example: data on customers without labels or target variables is required since clustering is an unsuperised machine learning technique). Therefore, one the right hand side, a general description of the needed input data can be downloaded by the non-technical user of a MCC and used to describe the data needs to the client organization. Further, there are multiple options to uplaod data fpr the model: either the non-technical user from the MCC already has access to the data and can upload it from local storage or from connected source (e.g. some cloud storage) or the user can create an upload link for their client so they can upload the data themselfes in a save manner.

REFERENCE	REQUIREMENT/ WORKFLOW-STEP
F1.2	The system must help users to identify the data needs of a use case to better define it.
F1.3	The system must be able to import training data from different sources.
F2.1	The system must enable non-technical users to explain the technology behind the chosen use case so it fosters trust in the outcomes.
F2.2	The system must allow users to adopt parameters in the execution of a use case according to their specific business needs.
F2.3	The system must allow the user to explain how data is kept and security of training data is provided to foster trust.
NF1.2	The proposal must guide users through an end-to-end workflow that begins with defining the use case to evaluating the first model.
NF 1.3	The proposal must display relevant explanations to the user at every step of the workflow so the user understands the process well enough to trust its results.
WF2	Define data needs and business value
WF3	Provide training data

Table 4: Requirements and Workflow-steps integrated in page 3 of the demonstration

Figure 9: Demonstration page 4

On page four of the demonstration the data is prepared for applying the clustering algorithm. The suer can explore the uplaoded data in a window on the left side of the interface. Below are recommendations on how to prepare the data. For example, outliers are automatically recognized and it is recommended to remove them. Further, columns with text values in them can be transformed to nummerical values so they are readable for the algorithm. Some AutoML technologies automatically test different options of tansforming the input features against performance metrics of the final model. This automated feature engineering and selection is harder to do in the case of unsupervised machine learning algorithms as it is hard to define a valid quality metric. So, in this case some manual transformations are better. Further, on the left side of the interface a user might want to enrich the input data. Appending more information to the input data will eventually improve the machine learning model. As an example, a customer dataset might contain addresses of customers. Using the zip code as a key a user could easily be offered to append publicly available information on their customer data (e.g. does the customer live in a city or a small village? Is it a rich or poor region?).
On the right side of the interace the user must select a parameter for the model. Usually, AutoML is used to optimize all parameters of applied algorithms. In the case of clustering the optimal number of clusters is highly dependent on the business need. Therefore, the tool gives a recommendation from a statistical point of view and a range in which the number of clusters should be chosen but lets the user have a final say in tuning this parameter. After preparing the data and defining the number of clusters, the user clicks on “RUN ANALYSIS”. Now, an existing AutoML technology is triggered in the backend of the tool (for example AutoML technologies of the hyperscaler aws or Google) and handles all the necessary steps to obtain a good result: feature selection, scaling nummeric values, selecting the optimal algorithm for the given data (in the case of clustering for example k-means, k-modes or DBSCAN), finetuning all the other paramters of the selcted algorithm, etc.

REFERENCE	REQUIREMENT/ WORKFLOW-STEP
F1.4	The system should help users to prepare training data to make them readable and useful for machine learning algorithms.
F.1.5	The system must be able to execute different AutoML technologies.
NF1.1	the proposal should never display information that is too technical or requires explicit knowledge on the process and methods of data science from its user
NF 1.3	the proposal must display relevant explanations to the user at every step of the workflow so the user understands the process well enough to trust its results.
WF4	Enrich data and handle problems
WF5	Start AutoML analysis

Table 5: Requirements and Workflow-steps integrated in page 4 of the demonstration

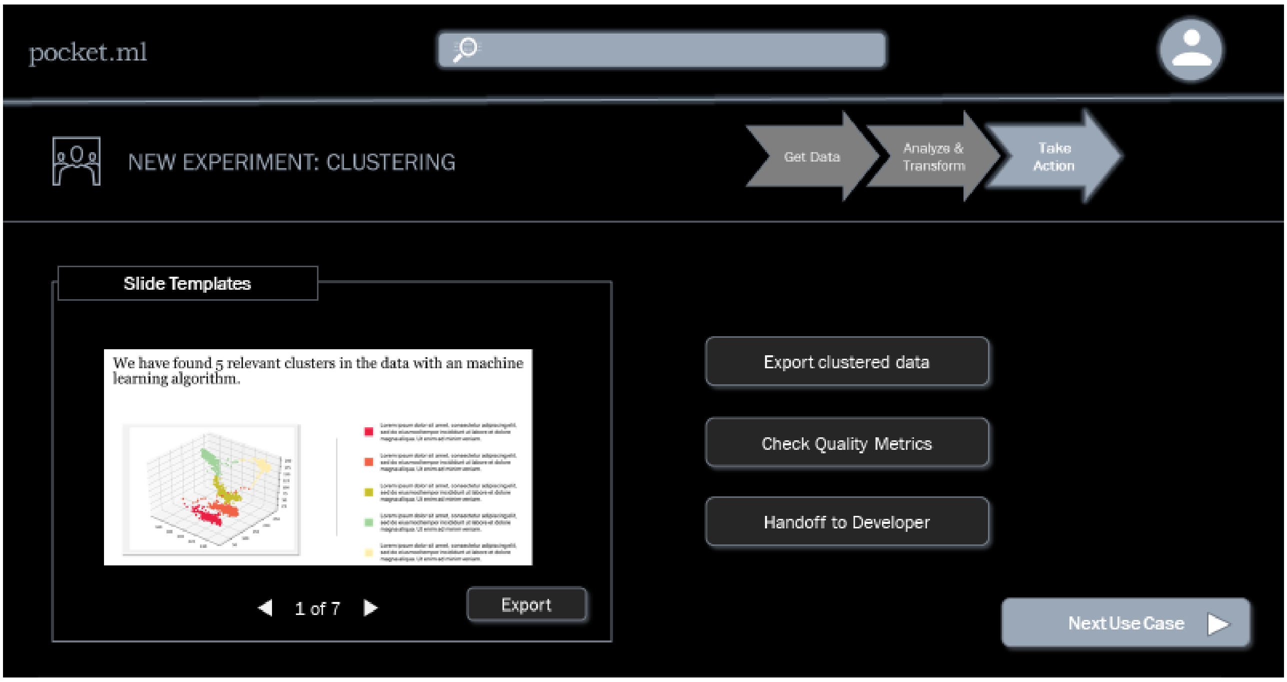


Figure 10: Demonstration page 5

Finally, on page five of the demonstration the results are presented to the user. The user can now export templates for slides specifically prepared for this use case and leverage them to present and explain the results of the analysis to their client. Further, the clustered data can be exported to use it in various ways. Quality metrics of the applied technology can be checked and if the use case is found to be valuable a handoff to an expert (developer) is made. The expert will review the analysis, append more data, create new features or better fit the model to the business context before preparing it for production.

REFERENCE	REQUIREMENT/ WORKFLOW-STEP
F2.4	The system must enable the user to explain the outputs of the AutoML system in a way that they can be translated to business value.
F3.1	The system must help machine learning experts to build production-ready models faster after a successful proof of concept.
NF1.1	the proposal should never display information that is too technical or requires explicit knowledge on the process and methods of data science from its user
NF1.2	The proposal must guide users through an end-to-end workflow that begins with defining the use case to evaluating the first model.
NF 1.3	the proposal must display relevant explanations to the user at every step of the workflow, so the user understands the process well enough to trust its results.
WF6	Interpret and present results to client
WF7	Make model production ready

Table 6: Requirements and Workflow-steps integrated in page 5 of the demonstration

5.5. EVALUATION

To evaluate my proposal, I used the demonstration explained in the chapter before and showed it to three experts, whom I each asked the same questions (compare chapter 2.2). The experts were selected based on their experience with machine learning techniques and the Management Consulting industry. Transcripts of the interviews can be found in "Appendix A: Interview Transcriptions". To guarantee anonymity for all interviewees I will refer to them with "Person 1", "Person 2", "Person 3". Person 1 works for one of the biggest Management Consulting Companies in the world as a Senior Manager and is responsible for a cloud-based data platform that integrates data and analytics tools in Germany. This platform currently tackles the strategic need for the Management Consulting company to incentivize and enable consultants that specialize in various industries to develop and sustain data-based assets. Person 2 holds a PhD in computer linguistics with a specialization in AI. They worked for one of the biggest Management Consulting Companies in the world until one month prior to the interview as an AI-Manager in a centralized hub of developers and machine learning experts. The function of said hub was to identify promising use cases for machine learning applications across the company and help develop and deploy such applications. That qualifies Person 2 perfectly to evaluate the proposal and the role that AutoML could play in changing the dynamics of such efforts to foster asset-based consulting in a large company. Now they work in a Start-up as Vice President of Data and AI. Person 3 holds a PhD in computer science with specialization in Data Science. They work for a medium sized management consultancy (500-1000 employees) that is specialized on digital solutions as a "Consultant for Data Science". It was important to not just include perspectives of the largest MCCs in the world but also have Person 3 that can evaluate the role AutoML can play in smaller consulting companies where business users and machine learning experts might not be that far away from each other – both in a physical and organizational sense.

All three interviews contained very positive feedback for the overall proposal which is the workflow depicted in chapter 5.3. Person 1 pointed out that the proposal would “fit right into our strategy going forward” as there is a big need to find scalable (asset-based) solutions to offer to their clients because it has become very hard to grow in their current business model. The current business model relies on hiring top talent and charging clients an hourly fee for their services like working out strategic priorities or improving business processes. The problem with that seems to be that (especially in countries like Germany) there is just not enough talent available to grow at the rates MCCs used to. Thus, MCCs are trying to reinvent their business models and sell asset-based subscriptions (for example for machine learning models and data-driven insights) as they rely less on labor. But Person 1 currently sees a problem in expertise concerning machine learning and AI being “scattered across the organization”, which makes it hard to identify and develop the most promising assets for their clients at an acceptable pace and quality. Therefore, it makes sense to enable business users to execute PoCs with AutoML and let experts focus on use cases where the value is already proven. Person 2 adds to that by pointing out that in their function at a central machine learning hub for another large MCC, “most ideas have not been tested before” and experts would have to spend a lot of time to understand the business context for the desired application and in assessing if it would even be advantageous to use machine learning techniques to solve a specific problem. Person 3 brings another angle into the discussion by highlighting that in their experience there is often a “misfit in […] between what business-oriented consultants sell and what technical experts can implement to achieve the biggest value for a client”. That also points toward the added value of helping business-users themselves test their ideas for machine learning applications by leveraging AutoML. After a PoC it becomes vastly easier to understand which value a machine learning model could bring in a certain application context and experts would only have to focus on use cases where there is an existing assessment about the feasibility of an idea.

Interestingly, many of the more critical comments from all interviewees focused rather on the interfaces of the demonstration than on the workflow itself. This goes to show that the proposal, which is the workflow and not the interfaces built to demonstrate the application of the workflow, achieved great acceptance, but also that for a successful implementation of the workflow, soft factors like user experience are key. In this evaluation I will focus mainly on the discussions around the workflow itself and not the design of the interfaces in the demonstration as that is where the study objectives are achieved.

There are two major patterns of the discussions with all interviewees. The first one is regarding the very first step in the workflow, namely defining the use case. Person 1 called this step of the workflow the “most valuable asset” in the whole proposal and liked the idea of assisting the user with a use case database in this step They added that in their company, great efforts are undertaken to foster the “build once, use anywhere principle”, meaning that use cases (i.e. machine learning models) could be copied from one client to another. The feedback was to not only think in use cases but also pretrained models that can be reused across client organizations with minor adaptations. Person 2 also saw this step as the greatest value-driver of the proposal and encouraged to think about even more possibilities to assist users in defining use cases. Person 2 said that just this step could be an own “workflow in itself”. Ideas to go beyond a use case database included implementing a chatbot that interacts with a user to better define a use case or to use project descriptions, that are usually created by consultants in the beginning of every project, in order to give recommendations for possible use cases based on historic data.

The second pattern in the discussions evolves around workflow-steps 2 and 3 (providing and preparing training data). Person 2 explained that “quality data is crucial in building models and most organizations have an overall bad data quality that is not ready for building AI models”. Person 3 added that users without expertise in machine learning might often face the problem that selected data is not fit for the purpose of building a desired machine learning model and asked “could a user for example just upload a dataset and see which use cases would be possible for the given dataset?”. Maybe users could even “get a probability of a successful model for a given use case and data that is uploaded” based on comparing datasets with existing implementations that were successful. A database for calculating such a probability of success could for example come from comparing datasets and use cases from machine learning platforms like “Kaggle.com”. Both points could indicate the need to add further steps in the workflow that ensure a certain data-quality is achieved and the selected data is fit-for-purpose.

6. DISCUSSION

Discussing the results from the evaluation of the proposal I want to place the most relevant topics that occurred into a greater context of existing literature and concepts. Then, I give an outlook on how these concepts could improve the proposal.

First, on the topic of the importance and challenges of defining use cases, it is interesting picking up on the idea mentioned by one of the interviewees to leverage a "build once, use anywhere" principle. That connects to concepts in existing literature like pre-trained machine learning models (PTMs) which work by "storing knowledge into huge parameters and fine-tuning on specific tasks" as described by Han et al. (2021, p. 225). They further elaborate that "It is now the consensus of the AI community to adopt PTMs as backbone for downstream tasks rather than learning models from scratch" (Han et al. 2021 p. 225). This is especially true in image recognition and NLP where big corporations such as aws[3] or NVIDIA[4] offer PTMs on online marketplaces to perform specific tasks like object recognition in images or summarizing texts. Further examples of a "build once, use anywhere" approach can be found in ideas to offer specific machine learning tasks as a reusable microservice (Pahl and Loipfinger 2018) or simply copying purpose-built machine learning classifiers (Unceta, Nin and Pujol 2020). The overall approach is somewhat comparable to the idea of using AutoML to enable business users to build machine learning models in that the heavy lifting concerning technical abilities is done in advance and not by the person applying the technology in any given business context. On the one hand, PTMs help adopting machine learning technologies by non-technical users because they define a very clear skill that can be performed, and users do not have to define use cases themselves. On the other hand, they offer less flexibility in building machine learning models as is possible by leveraging AutoML technologies to build custom models for the business context of each desired application. If PTMs would offer the same kind of flexibility, we would be very close to achieving Artificial General Intelligence (AGI), which still seems to be far away - and where researchers are not on one page whether this is something that is unconditionally to be desired (Fjelland 2020). Summarizing on this topic, it might be a useful add-on to offer a marketplace-like user experience where management consultants could choose from PTMs for specific tasks in the proposal. But this idea shouldn't replace the approach to use AutoML to build custom machine learning applications for specific problems of each client as that would rob MCCs of one of their biggest strengths in client relationships, namely their expertise on specific problems of their clients and creativity to solve these problems. Examples of user experiences that successfully merge marketplace-characteristics and the freedom to build custom models exist in tools like KNIME Hub, where users of the no-code machine learning tool can share workflows of the visual programming language that is also used to build custom models (Ordenes and Silipo 2021).

Second, on the topic of the provision and quality of available training data within client organizations we enter the research field of data-centric AI. In an article for Harvard Business Review, Andrew Ng, Adjunct Professor at Stanford University's Computer Science Department and (co-)founder of multiple successful AI and educational companies, states that for AI to fulfil its potential in industries

[3] https://aws.amazon.com/marketplace/solutions/machine-learning/pre-trained-models/
Last accessed: 10.10.2020
[4] https://developer.nvidia.com/ai-models
Last accessed: 10.10.2020

outside of big consumer internet companies “we need teams that can program with data rather than program with code” (Ng 2021). What is meant by that is that the bottleneck for most successful AI (and therefore machine learning) adoptions is not the algorithms or tools to help you develop such algorithms available but the quality of the data that is used to train a model. Thus, other authors conceptualize data as a production factor and highlight its strategic role in gaining a competitive advantage (Huang et al 2021). A promise of focusing on data in advancing efforts to leverage machine learning or AI techniques is that knowledge of domain experts can be better captured when including them in preparing the data for such models. If it is too laborious for a domain expert to label enough data to train a machine learning model, there are several techniques in data-centric AI to help create labelled data more efficiently. Techniques worth mentioning in this context include active learning, where a domain expert only labels a limited but expressive sample of the original dataset which is used to train an algorithm (Settles, Craven and Ray 2007), and weak supervision, where a training function is used to translate expertise on the subject at hand into labels for training datasets (Ratner et a. 2020). Further, there is also the technique of confident learning to improve the quality of labels in training datasets (Northcott, Jiang and Chuang 2021). Data-centric AI techniques like the ones mentioned can be perfectly combined with AutoML as they together offer the advantage of capturing a domain expert’s knowledge in preparing high-quality data while not requiring skills in coding to develop models. With the project “The automated statistician”, there is also an approach that can be counted as data-centric AI as it helps to explore training data properties and their influence on machine learning models, that uses AutoML techniques (Steinruecken et al. 2019. Summarizing on this topic, for the proposal given in this study it might be worthwhile to include workflow-steps that allow users to easily label unlabelled data, explore and improve the quality of training data. It needs to be evaluated whether a business user can conduct these tasks or if experts must be involved here.

7. CONCLUSIONS

7.1. SYNTHESIS OF THE DEVELOPED WORK

"We will need to find more scalable solutions to grow in the future. Our talent concerning machine learning and AI is scattered across the organization and it's hard to connect them." This is a quote from the interview with Person 1 that sticked with me and made all the efforts put into this thesis worthwhile. MCCs have a problem. They need to get smarter, faster and more creative about developing more scalable services like machine learning models to solve their client's problems. They cannot solve this problem just by hiring more and more technology experts and developers. It is an organizational problem. Not only MCCs face this organizational problem, but I am also convinced it is at least as prevalent in most other industries and even worse in industries that have a lesser ability to hire top talent than MCCs. AutoML can help mitigate this problem. If used in the right way, it can be a powerful tool in the hands of business users to quickly validate their ideas and get experts involved where they are most needed. Then, AutoML helps these experts to build the most powerful models for the problem at hand. This study explored how to do that. In doing that, it also shined a light on how important it is not to just provide the technology but also to help users define use cases, reuse ideas and prepare data in a way that machine learning can perform best on them.

Starting this research endeavor, I set specific objectives that were to be reached in this study. First, a diagnosis of companies needs was to be conducted. In chapter 4, I conducted a literature review on the Management consulting industry and summarized its results in a SWOT analysis. Major learnings were that MCCs need to find more scalable, subscription-based solutions to grow at the same rates that they are used. This was later confirmed in the expert-interviews. Second, areas and tasks suitable for applying AutoML tools were to be identified. In chapter 3, I conducted a literature review on (automated) machine learning and their business applications. The major learning was that AutoML can democratize building ML models but is best used to build PoCs as trust is elementary for ML outputs to be transformed in business value and trust might only come with experts being involved at the right place. Together with the analysis on the Management Consulting industry, I was able to build a theory as a first answer to the research question and could derive requirements for an information system from that theory (chapters 5.1 and 5.2). Third, I built a proposal in the form of a workflow that shows how an AutoML tool can help organizations (here: specifically MCCs) and their non-technical users to use machine learning techniques (chapter 5.3). Fourth, I built a demonstration for the workflow that also includes non-functional requirements (chapter 5.4). Finally, the proposal was evaluated by experts (chapter 5.5) and ideas for improving the proposal were discussed in chapter 6.

7.2. LIMITATIONS AND FUTURE WORK

Although the research objectives were reached and contributions to academic literature as well as the world of practitioners were made, this study has some limitations. First and foremost, it was a big learning that the design of user interfaces for an information system like the one described with the workflow (proposal) is crucial for its acceptance. That became obvious in the evaluation where some of the feedback was on the proposal itself while other feedback was mainly about its demonstration. The design of user interfaces was not the main object of analysis in this study, which is why future research will need to dive deeper into this topic. Researchers could build on first interesting approaches to design the interaction of humans with AI or ML systems (Subramonyam, Im, Seifert & Adar 2022 and Ostheimer, Chowdhury & Iqbal 2021) but my recommendation is to approach the topic in a more experimental way. One approach could be to conduct a high number of interviews with practitioners on different designs to learn which aspects inform user acceptance in which way.

Further, one of the improvement ideas for the proposal was to focus more on the step of finding and defining a use case. In this study I was not able to give a final answer on which representation would help users best to define their use case. This is a question that could also be tackled with experimental research methods like A/B testing (Kohavi and Longbotham 2017). The same is to be said about the question that was discussed of which methods would work best for business users and/ or experts to help client organizations improve the data quality for building ML models.

Lastly, in trying to find a clear analytical lens, this study was focused on the Management consulting industry. It is not yet clear as to how well the learnings of this study can be transferred to other industries. Future research could consist of evaluating the proposal given in this study in different contexts. However, researchers will also need to evaluate per industry whether there is a AutoML tool that already fulfills the requirements of said industry. In the case of MCCs chapter 3.3.2 and the evaluations showed that none such tool exists on the market at the moment but that many of the functionalities of the desired information system are already implemented by different organizations and would need to be plugged together for a comprehensive workflow.

BIBLIOGRAPHY

Alshemali, B. and Kalita,J. (2020). Improving the Reliability of Deep Neural Networks in NLP: A Review. Knowledge-Based Systems, Volume 191, 2020, 105210, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2019.105210.

Asamoah, Daniel & Sharda, Ramesh & Zadeh, Amir & Kalgotra, Pankush. (2017). Preparing a Data Scientist: A Pedagogic Experience in Designing a Big Data Analytics Course: Preparing a Data Scientist. Decision Sciences Journal of Innovative Education. 15. 161-190. 10.1111/dsji.12125.

Belle V, Papantonis I (2021). Principles and Practice of Explainable Machine Learning. Front Big Data. 2021 Jul 1;4:688969. doi: 10.3389/fdata.2021.688969. PMID: 34278297; PMCID: PMC8281957.

Brent Kitchens, David Dobolyi, Jingjing Li & Ahmed Abbasi (2018). Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data. Journal of Management Information Systems, 35:2, 540-574, DOI: 10.1080/07421222.2018.1451957

Borges, A, Laurindo, F., Spínola, M., Goncalves, R., Mattos, C. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. International Journal of Information Management 57 (2021) 102225. https://doi.org/10.1016/j.ijinfomgt.2020.102225

Bose, R. (2009), "Advanced analytics: opportunities and challenges", Industrial Management & Data Systems, Vol. 109 No. 2, pp. 155-172. https://doi.org/10.1108/02635570910930073

Blohm, Matthias & Hanussek, Marc & Kintz, Maximilien. (2021). Leveraging Automated Machine Learning for Text Classification: Evaluation of AutoML Tools and Comparison with Human Performance. 1131-1136. 10.5220/0010331411311136.

Brocke, Jan vom & Hevner, Alan & Maedche, Alexander. (2020). Introduction to Design Science Research. in: Jan vom Brocke & Alan Hevner & Alexander Maedche (ed.), Design Science Research. Cases, pages 1-13. Springer.10.1007/978-3-030-46781-4_1.

Brändle, Sebastian & Hanussek, Marc & Blohm, Matthias & Kintz, Maximilien. (2021). Evaluation of Representation Models for Text Classification with AutoML Tools.

Canback, Staffan (1998). The Logic of Management Consulting, Part 1. Journal of Management Consulting, Vol. 10, No. 2, pp. 3-11, 1998, Available at SSRN: ttps://ssrn.com/abstract=1276066

Christensen CM, Wang D, van Bever D (2013) Consulting on the cusp of disruption. Harvard Business Review 91(10):106–114

Coleman, J. S. (1990). Foundations of Social Theory. Cambridge, MA: Belknap Press of Harvard University Press

Crisan, A. and Fiore-Gartland, B. (2021). Fits and Starts: Enterprise Use of AutoML and the Role of Humans in the Loop. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 601, 1–15. https://doi.org/10.1145/3411764.3445775

De Massis, Alfredo and Foss, Nicolai (2018). “Advancing Family Business Research: The Promise of Microfoundations”. Family Business Review 31/2018, p. 386 – 396. DOI: 10.1177/0894486518803422

Edwards, J (2016). Management consulting in the US. IBISWorld Inc, Los Angeles.

Elsken,T, Metzen, J. and Hutter, F. (2019). Neural Architecture Search. In: Hutter, F., Kotthoff, L., Vanschoren, J. (eds) Automated Machine Learning. The Springer Series on Challenges in Machine Learning. Springer, Cham. https://doi.org/10.1007/978-3-030-05318-5_9

European Federation of Management Consultancies Associations: FEACO (2021). The Management Consulting Industry. Retreived on 27 January 2022, from https://www.feaco.org/industry

Federation of German Industries, BDI (24.02.2021). Datenwirtschaft in Deutschland. Wo stehen die Unternehmen in der Datennutzung und was sind ihre größten Hemmnisse? BDI e.V., Germany. https://bdi.eu/publikation/news/datenwirtschaft-in-deutschland/

Flynn, M. and Kowalkiewicz, M. (2018). Opportunities and Risks of Digital Business Model Innovation for Behemoths in Consulting. In: Nissen, V. (2018). Digital Transformation of the Consulting Industry. Extending the traditional delivery model. Springer International Publishsing AG, Switzerland.

Fjelland, R. Why general artificial intelligence will not be realized (2020). Humanit Soc Sci Commun 7, 10 (2020). https://doi.org/10.1057/s41599-020-0494-4

German Chancellery (2021). Datenstrategie der Bundesregierung – Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum. Kabinettfassung, 27. Januar 2021. Bundeskanzleramt, Berlin. https://www.bundesregierung.de/resource/blob/992814/1845634/f073096a398e59573c7526feaadd43c4/datenstrategie-der-bundesregierung-download-bpa-data.pdf?download=1

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal*. https://doi.org/10.1038/bdj.2008.192

Global Management Consultants Industry - Market Research Report. (2021, January 30). IBISWorld. Retreived on 27. January 2022, from https://www.ibisworld.com/global/market-research-reports/global-management-consultants-industry/

Goetz, M., Leganza, G., Granzen and A., Hennig, C. (2021 a). Now Tech: AI Consultancies, Q1 2021. Forrester's Overview Of 40 AI Service Providers. Forrester February 4, 2021.

Goetz, M., Leganza, G., Granzen and A., Hennig, C. (2021 b). The Forrester Wave™: AI Consultancies, Q1 2021. The 13 Providers That Matter Most And How They Stack Up. Forrester March 25, 2021.

Granzen, A., Michele, G., Giron, F., Leganza, G., Lin, N., Hennig, C. and Nagel, B. (2019). Consultancies Are Reinventing Their Service Model For AI. Insights From The Forrester Wave™ Evaluations Of AI Consultancies. Forrester, December 19, 2019.

Greiner L, Metzger R (1983). Consulting to management. Prentice-Hall, Englewood Cliffs.

Gupta, M. and George, J. F. (2016): Toward the development of a big data analytics capability. Information & Management, Volume 53, Issue 8, December 2016, Pages 1049-1064 .

Han, X., Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, Yuqi Huo, Jiezhong Qiu, Yuan Yao, Ao Zhang, Liang Zhang, Wentao Han, Minlie Huang, Qin Jin, Yanyan Lan, Yang Liu, Zhiyuan Liu, Zhiwu Lu, Xipeng Qiu, Ruihua Song, Jie Tang, Ji-Rong Wen, Jinhui Yuan, Wayne Xin Zhao, Jun Zhu (2021). Pre-trained models: Past, present and future, AI Open, Volume 2, 2021, Pages 225-250, ISSN 2666-6510, https://doi.org/10.1016/j.aiopen.2021.08.002.

Harari, Y (2017). Reboot for the AI revolution. Nature 550, 324–327 (2017). https://doi.org/10.1038/550324a

Hardt, C. (2018). The Best of Two Worlds—Digitization of Matchmaking Between Consulting Firms and Independent Consultants. In: Nissen, V. (2018). Digital Transformation of the Consulting Industry. Extending the traditional delivery model. Springer International Publishsing AG, Switzerland. 10.1007/978-3-319-70491-3_1.

Helms, M.M. and Nixon, J. (2010). Exploring SWOT analysis – where are we now? A review of academic research from the last decade. Journal of Strategy and Management, Vol. 3 No. 3, pp. 215-251. https://doi.org/10.1108/17554251011064837

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly: Management Information Systems*. https://doi.org/10.2307/25148625

Ho, T.K. (2002). Multiple classifier combination: lessons and next steps. Hybrid Methods in Pattern Recognition (2002), pp. 171-198, 10.1142/9789812778147_0007

Huang, L., Yifan Dou, Yezheng Liu, Jinzhao Wang, Gang Chen, Xiaoyang Zhang, Runyin Wang (2021). Toward a research framework to conceptualize data as a factor of production: The data marketplace perspective. Fundamental Research, Volume 1, Issue 5, 2021, Pages 586-594, ISSN 2667-3258, https://doi.org/10.1016/j.fmre.2021.08.006.

Hutter, F., Kotthoff, L. and Vanschoren, J. (2019). Automated Machine Learning – Methods, Systems, Challenges. In: The Springer Series on Challenges in Machine Learning. Springer International Publishing. DOI: 10.1007/978-3-030-05318-5

Jarvis, D (2020). The AI talent shortage isn't over yet. Deloitte Insights, September 30, 2020. Retreived on 27. January 2022, from https://www2.deloitte.com/us/en/insights/industry/technology/ai-talent-challenges-shortage.html

Jerónimo, C., Pereira, L. and Sousa, H. (2019). Management Consulting Business Models: Operations through and for Digital Transformation. *IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2019, pp. 1-6, doi: 10.1109/ICE.2019.8792592.

Kennet, R.S. and Redman, T.C. (2019): The Real Work of Data Science. Turning Data into Information, Better Decisions and Stronger Organizations. Wiley. ISBN: 978-1-119-57070

Kohavi, Ron & Longbotham, Roger. (2017). Online Controlled Experiments and A/B Testing. 10.1007/978-1-4899-7687-1_891.

Li, Y., Chu, X., Tian, D., Feng, J., Mu, W. (2021). Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm. Applied Soft Computing, Volume 113, Part B, 2021, 107924, ISSN 1568-4946. https://doi.org/10.1016/j.asoc.2021.107924.

Lee, I., Shin, Y.J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. Business Horizons (2020) 63, p. 157-170.

Maleki, F., Ovens, K., Najafian, K., Forghani, B, Reinhold, C., Forghani, R. Overview of Machine Learning Part 1: Fundamentals and Classic Approaches. Neuroimaging Clinics of North America, Volume 30, Issue 4, 2020, Pages e17-e32. https://doi.org/10.1016/j.nic.2020.08.007.

Management Consultants in Europe - Market Research Report. (2021, January 15). IBISWorld. Retreived on 27. January 2022, from https://www.ibisworld.com/eu/industry/management-consultants/3905/

Mathias Kraus, Stefan Feuerriegel, Asil Oztekin (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications, European Journal of Operational Research, Volume 281, Issue 3, 2020, Pages 628-641, ISSN 0377-2217, https://doi.org/10.1016/j.ejor.2019.09.018.

March, Salvatore & Smith, Gerald. (1995). Design and Natural Science Research on Information Technology. Decision Support Systems. 15. 251-266. 10.1016/0167-9236(94)00041-2.

McKenna, C. D. (1995), 'The Origins of Modern Management Consulting', Business and Economic History, 24(1), 51–8.

Mäkinen, S., Skogström, H., Laaksonen, E., Mikkonen, T. (2021). Who Needs MLOps: What Data Scientists Seek to Accomplish and How Can MLOps Help? 2021 IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN) of 43rd International Conference on Software Engineering (ICSE)
arXiv:2103.08942

Nissen, Volker. (2018). Digital Transformation of the Consulting Industry—Introduction and Overview. In: Nissen, V. (2018). Digital Transformation of the Consulting Industry. Extending the traditional delivery model. Springer International Publishsing AG, Switzerland. 10.1007/978-3-319-70491-3_1.

Ng, Andrew (2021). AI Doesn't Have to Be Too Complicated or Expensive for Your Business. Harvard Business Review online 29.07.2021. https://hbr.org/2021/07/ai-doesnt-have-to-be-too-complicated-or-expensive-for-your-business

Northcutt, C., Jiang, L. and Chuang, I. (2021). Confident learning: Estimating uncertainty in dataset labels. Journal of Artificial Intelligence Research, 70, pp.1373-1411.

Olschewski, S., Diao, L. and Rieskamp, J. (2021). Reinforcement learning about asset variability and correlation in repeated portfolio decisions, Journal of Behavioral and Experimental Finance, Volume 32, 2021, 100559, ISSN 2214-6350, https://doi.org/10.1016/j.jbef.2021.100559.

Ordenes, F., Silipo, R. (2021). Machine learning for marketing on the KNIME Hub: The development of a live repository for marketing applications, Journal of Business Research, Volume 137, 2021, Pages 393-410, ISSN 0148-2963 https://doi.org/10.1016/j.jbusres.2021.08.0

Ostheimer, J., Chowdhury, S. and Iqbal, S. (2021). An alliance of humans and machines for machine learning: Hybrid intelligent systems and their design principles. Technology in Society, Volume 66, 2021, 101647, ISSN 0160-791X.

Pahl, M.-O. and Loipfinger, M.(2018). Machine learning as a reusable microservice, NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium, 2018, pp. 1-7, doi: 10.1109/NOMS.2018.8406165.

Peffers, K., Tuunanen, T., Rothenberger, M.A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24 (3), 45-77.

Poulakis, Yannis & Doulkeridis, Christos & Kyriazis, Dimosthenis. (2020). AutoClust: A Framework for Automated Clustering Based on Cluster Validity Indices. 1220-1225. 10.1109/ICDM50108.2020.00153.

Pressman, Roger (2009). Software Engineering: A Practitioner's Approach. 7th edition , McGraw Hill, ISBN 10: 9780078022128

Reilly, A., Depa, J., Douglass, G. (2019). AI: BUILT TO SCALE. From experimental to exponential. Accenture.
https://www.accenture.com/us-en/insights/artificial-intelligence/ai-investments#:~:text=Companies%20in%20our%20study%20that,three%20key%20financial%20valuation%20metrics.
last accessed: August 8th 2022, 17 h

Ruf, P.; Madan, M.; Reich, C.; Ould-Abdeslam, D (2021). Demystifying MLOps and Presenting a Recipe for the Selection of Open-Source Tools. Appl. Sci. 2021, 11, 8861. https://doi.org/10.3390/app11198861

Sagi, Omer & Rokach, Lior. (2018). Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 8. e1249. 10.1002/widm.1249.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J. and K. -R. Müller (2021) "Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications," in Proceedings of the IEEE, vol. 109, no. 3, pp. 247-278, March 2021, doi: 10.1109/JPROC.2021.3060483

Seeber,I., Bittner, E., Briggs, R., de Vreede, T., de Vreede, G., Elkins, A., Maier, R., Merz, A. Oeste-Reiß, S., Randrup, N., Schwabe, G. and Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. Information & Management, Volume 57, Issue 2, 2020, 103174, ISSN 0378-7206, https://doi.org/10.1016/j.im.2019.103174.

Settles, B., Craven, M. and Ray, S. (2007). Multiple-instance active learning. Advances in Neural Information Processing Systems, 20, pp.1289-1296.

Schumpeter, Joseph A. (1950). Capitalism, Socialism and Democracy. 3rd ed. New York: Harper-Collins.

Schüritz, Ronny; Brand, Ella; Satzger, Gerhard; and Bischhoffshausen, Johannes, (2017). HOW TO CULTIVATE ANALYTICS CAPABILITIES WITHIN AN ORGANIZATION? – DESIGN AND TYPES OF ANALYTICS COMPETENCY CENTERS. In Proceedings of the 25th European Conference on Information Systems (ECIS), Guimarães, Portugal, June 5-10, 2017 (pp. 389-404). ISBN 978-989-20-7655-3 Research Papers. https://aisel.aisnet.org/ecis2017_rp/26

Sculley, D.; Holt, G.; Golovin, D.; Davydov, E.; Phillips, T.; Ebner, D.; Chaudhary, V.; Young, M.; Crespo, J.F.; Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. In Advances in Neural Information Processing Systems; Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc.: Cambridge, MA, USA, 2015; Volume 28.

Shin, D: (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI, International Journal of Human-Computer Studies, Volume 146, 2021, 102551, ISSN 1071-5819, https://doi.org/10.1016/j.ijhcs.2020.102551.

Shubhra Kanti Karmaker ("Santu"), Md. Mahadi Hassan, Micah J. Smith, Lei Xu, Chengxiang Zhai, and Kalyan Veeramachaneni. 2021. AutoML to Date and Beyond: Challenges and Opportunities. ACM Comput. Surv. 54, 8, Article 175 (November 2022), 36 pages. DOI:https://doi.org/10.1145/3470918

Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, Lanctot M, Sifre L, Kumaran D, Graepel T, Lillicrap T, Simonyan K, Hassabis D (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science. 2018 Dec 7;362(6419):1140-1144. doi: 10.1126/science.aar6404.

Sommerville, Ian (2010). Software Engineering. 9th Edition, Pearson Education Inc., publishing as Addison-Wesley.

Song, R., Jang, S., Wang, Y., Hanssens, D. and Suh, J. (2021). Reinforcement learning and risk preference in equity linked notes markets. Journal of Empirical Finance, Volume 64, 2021, Pages 224-246, ISSN 0927-5398, https://doi.org/10.1016/j.jempfin.2021.09.004.

Steinruecken, C., Smith, E., Janz, D., Lloyd, J., Ghahramani, Z. (2019). The Automatic Statistician. In: Hutter, F., Kotthoff, L., Vanschoren, J. (eds) Automated Machine Learning. The Springer Series on Challenges in Machine Learning. Springer, Cham. https://doi.org/10.1007/978-3-030-05318-5_9

Subramonyam, H. Im, H., Seifert, C. and Eytan Adar (2022). Human-AI Guidelines in Practice: The Power of Leaky Abstractions in Cross-Disciplinary Teams. In CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022,

New Orleans, LA, USA. ACM, New York, NY, USA, 31 pages. https://doi.org/10.1145/3491102.3517537

Tseng, K., Lin, J., Chen, C. and Hassan,M. (2021). A fast instance segmentation with one-stage multi-task deep neural network for autonomous driving, Computers & Electrical Engineering, Volume 93, 2021, 107194, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2021.107194

Turner, A. N. (1982). Consulting is more than giving advice. Harvard Business Review 60 (September-October): 120–129.

Unceta, I., Nin, J. and Pujol, O.(2020). Copying Machine Learning Classifiers, in IEEE Access, vol. 8, pp. 160268-160284, 2020, doi: 10.1109/ACCESS.2020.3020638.

Werth, D. and Greff, T. (2018). Scalability in Consulting: Insights into the Scaling Capabilities of Business Models by Digital Technologies in Consulting Industry. In: Nissen, V. (2018). Digital Transformation of the Consulting Industry. Extending the traditional delivery model. Springer International Publishsing AG, Switzerland.

Wever, M., Tornede, A., Mohr, F., Hullermeier, E. (2021). AutoML for Multi-Label Classification: Overview and Empirical Evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. 1-1. 10.1109/TPAMI.2021.3051276.

Yuheng Hu, Anbang Xu, Yili Hong, David Gal, Vibha Sinha & Rama Akkiraju (2019). Generating Business Intelligence Through Social Media Analytics: Measuring Brand Personality with Consumer-, Employee-, and Firm-Generated Content, Journal of Management Information Systems, 36:3, 893-930, DOI: 10.1080/07421222.2019.1628908

APPENDIX A: INTERVIEW TRANSCRIPTIONS

Interview 1

Name (for reference): Person 1

Role and Background: Works for one of the biggest Management Consulting Companies in the world as a Senior Manager and is responsible for the cloud-based platform where consultants can work with data and analytics tools in Germany.

Answers to the questions:

1. Q: What is your general impression of the proposal?
   A: The proposal fits right into our strategy going forward. Asset-based consulting will play a huge role for us in the future. Our problem currently is that we cannot grow significantly anymore using our old business model. There are just not enough qualified people to hire anymore. There is some movement between consulting firms but no significant new workforce. We will need to find more scalable solutions to grow in the future. Our talent concerning machine learning and AI is scattered across the organization and it's hard to connect them. It makes sense to enable business users to build PoCs to find the cases where it is actually worth investing.

2. Q: Do you think you/ your organization would use the proposal if it was a real product?
   A: Our central data platform develops in the same direction. As I said, the idea fits right into our current strategy. The most valuable asset might be the use case database. It would be necessary to convince people to try it the tool initially though and there is a lot of communication involved.

3. Q: What are your recommendations to improve the proposal?
   A: There is even more tech-language to not display in the proposal (e.g. clustering or anomaly detection). Most people don't know which techniques they want to use but they care about what outcomes they can produce. Highlight the subtitles (questions) you used for describing the use case as they describe what users actually care about. It would be valuable to foster the "build once, use anywhere" principle. For that you would need more specific use cases so people get inspired. Then consultants could just copy models. Concerning UX, I would recommend using targeted suggestions for use cases depending on the profile of the user and an intelligent search engine to find use cases according to their tags. Intelligent search is better than a lot of filtering options.

Interview 2

Name (for reference): Person 2

Role and Background: Has a PhD in computer linguistics. Worked for one of the biggest Management Consulting Companies in the world until one month prior to the interview as an AI-Manager in a centralized hub of developers and machine learning experts. Now works in a Start-up as Vice President of Data and AI.

Answers to the questions:

1. Q: What is your general impression of the proposal?
   A: I like the overall proposal and it would be super valuable to have a workflow that helps finding the most promising applications for machine learning in an organization so experts can focus on the stuff with great business value. For many ideas that came to our hub when I was working in management consulting, machine learning or AI was not even the best possible solution to the problem. These ideas were not tested before, and experts had to carefully assess the application of such ideas before starting to develop solutions. The strength of the proposal is definitely in the beginning of the workflow where it is a very important part to assist the user in defining a use case. Maybe this is even a whole workflow in itself where a user is led to find a use case by answering different questions. Here you cannot do enough to really understand what users actually want to do and to build their expectations about what is possible.

2. Q: Do you think you/ your organization would use the proposal if it was a real product?
   A: The workflow-component helps non-technical users to execute a first PoC, I really like that. I think some of the functionalities of the proposal might best be delivered by integrating to existing tools. For example, when doing data preparation, Alteryx is a tool that is widely used in the consulting world. Having the overall workflow on top of it is very helpful though because it is hard for non-technical users to know which sequential steps to take when doing a PoC and just arranging which steps to take and which tools can be used for that is very helpful. I am a bit skeptical towards the quality of input data and if it can be easily transformed and used for building PoCs without an expert in data science or machine learning. Preparing the data is often one of the most creative and challenging tasks in data science. I don't know how well AutoML performs in automating these tasks. Preparing quality data is crucial in building models and most organizations have an overall bad data quality that is not ready for building AI models.

3. Q: What are your recommendations to improve the proposal?
   A: From a User Experience perspective it is interesting to think about different ideas on how to help users define a use case. Maybe a chatbot or something similar could be a viable solution to help users find the right specific use case for their problem? Also, it would be very cool if use cases could be recommended according to the project a consultant works on. Maybe one can achieve this trough analyzing project descriptions. Further, there should always be the possibility to call an expert in every step of the workflow.

Interview 3

Name (for reference): Person 3

Role and Background: Has a PhD in computer science with specialization in Data Science. Works for a medium sized management consultancy (500-1000 employees) that is specialized on digital solutions as a "Consultant for Data Science".

Answers to the questions:

1. Q: What is your general impression of the proposal?
   A: In my experience there is often a misfit in consulting between what business-oriented consultants sell and what technical experts can implement to achieve the biggest value for a client. Having a workflow where business-oriented consultants build small PoCs themselves might help to mitigate that misfit as a clearer picture emerges as to how and where machine learning techniques should be applied to create value for a client before technical experts get involved. Further,

2. Q: Do you think you/ your organization would use the proposal if it was a real product?
   A: The most important step is to understand what actually should happen on the business side. Also, the feasibility of what one wants to do with regards to data availability and meaningfulness for the problem at hand is a huge problem. To use the proposal, it must further address these aspects. One aspect of focusing on improving the data for a PoC could also be to have an extra step for debiasing training data. Machine learning platforms like AzureML have quite comprehensive functionalities there. Further, synthesizing training data to comply with the GDPR would be an interesting functionality.

3. Q: What are your recommendations to improve the proposal?
   A: To address the problem of data quality and meaningfulness it would be helpful to have an extra step where the tool could assess data quality and if it is fit for the purpose if the selected use case by predefined criteria per use case. Could a user for example just upload a dataset and see which use cases would be possible for the given dataset? It would be amazing to get a probability of a successful model for a given use case and data that is uploaded by the user. Maybe one could calculate this probability by analyzing past projects. After the analysis it might be an idea to rather offer dashboards integrated in existing BI systems instead of slides. Dashboards are usually better to present results of machine learning models and they could be continuously updated when the models change.