Fakultät Wirtschaftswissenschaften Lehrstuhl für Wirtschaftsinformatik | Business Intelligence Research

DATA SCIENCE AND ANALYTICS IN INDUSTRIAL MAINTENANCE:

SELECTION, EVALUATION, AND APPLICATION OF DATA-DRIVEN METHODS

DISSERTATION

to achieve the academic degree
Doctor rerum politicarum (Dr. rer. pol.)

by
Patrick Zschech, M.Sc.
born on June 24, 1988 in Dresden

First reviewer:	Prof. Dr. Susanne Strahringer
Second reviewer:	Prof. Dr. Udo Buscher

Date of submission:	June 8, 2020
Date of defense:	August 25, 2020

Acknowledgements

The marathon project of writing a doctoral thesis is not an individual runner's achievement. Instead, it is a team effort with many collaborators who either train ambitiously with you to reach the same goal, remove insurmountable hurdles on a bumpy path, or simply accompany you for a while to turn a tough race into an enjoyable journey. So I would like to take this opportunity to thank at least some of those who have supported me over the past few years in one way or another to complete this life-changing project.

First of all, I would like to express my sincere gratitude to my doctoral advisor, Prof. Andreas Hilbert (†). Since my early days at the TU Dresden, he inspired me as a passionate lecturer and aroused in me a fascination for the field of business analytics. Despite his long and debilitating illness, he supported my research ideas and consistently pushed me to improve my work. He taught me what is essential in life to maintain a careful balance between professional fulfillment and quality time with one's family. Your tireless spirit will always burn in my heart to remind me what is worth fighting for. May your soul rest in peace.

I would also like to express my deepest appreciation to my co-advisor, Prof. Susanne Strahringer. Without her support, these lines would likely not have been written. When I struggled, she found the right words to motivate me and keep me from giving up, while her constructive advice and her honest feedback have broadened my horizon more than once. During this entire journey, there was no concern to which she could not immediately respond with an adequate solution or a useful reference for further help. Thank you for giving me your valuable time and for encouraging me with your faith in my work and my abilities.

Furthermore, I wish to thank several more faculty members for their support and invaluable advice. Special thanks to Prof. Udo Buscher, who, without any hesitation, stepped in as an additional supervisor at short notice. I cannot express how grateful I am for your willingness to support me in the rapid completion of my doctoral project to meet the requirements for the next academic milestone. I also want to thank Prof. Werner Esswein for being my toughest and, likewise, most honest critic, as well as Prof. Eric Schoop, for his constantly encouraging advices. I have appreciated every discussion with each of you over the past twelve years, and I have enjoyed your outstanding instruction during my studies in the Bachelor and Master program. I also want to extend my gratitude to Prof. Christian Janiesch, who worked as an interim professor at our chair during the final year of my doctorate. During this short period, I learned so much from his broad scientific experience and professional knowledge, which he was always happy to share with no compensation. Thank you for your guidance along this amazing journey. I look forward to any opportunities that will bring us together in the future.

Without a doubt, the journey would not have been as pleasurable without my highly respected colleagues and peers at the chair. Thus, I want to thank Uwe Wieland for paving my academic path and being an amazing supervisor during my Bachelor and Master studies, Marcus Pfitzner for our trustful relationship that helped us conquer any kind of concern in challenging times, Matthias Lohse and Marco Barthel for ensuring a reliable IT infrastructure and many entertaining discussions during our lunch breaks, Conny Schumann, Karolin Stefani, and Michael Seifert for

creating a stimulating working environment to organize and run all our teaching and research activities, and Kerstin Petzold for making everyone feel welcomed and comfortable in the office. Furthermore, I want to express my particular appreciation to my research buddy, Kai Heinrich. Despite rough times, we managed to keep things together and come up with a great concept to meet all our academic obligations. Thank you for all the intensive scientific discussions, the joint development of teaching arrangements, the mutual pushing for publications, and our marvelous shared sense of humor. With pride, I look back at our achievements over the past few years, and with joy, I look forward to everything that will follow.

Beyond all academic contributors, I am deeply indebted to my family and friends. Above all, I want to thank my wife, who always gives me her full support in realizing my dreams, this doctoral thesis in particular. Andrea, you have been there from the very beginning, and together, we have overcome all the ups and downs. In times of setbacks, you stood by my side and held my hand, and in times of success, you shared and enjoyed my happiness. Thank you for running this race with me. I am so grateful to call you my teammate, my companion, my best advisor, and my partner in crime. I also want to thank my parents and friends for their confidence, patience, and understanding throughout these years. Without your backing, it would not have been possible to realize this project. Finally, I cannot close without acknowledging the intense energy I have experienced through my little son, Sebastian. With his wild, curious, witty, and wonderful nature, he kept me running the full distance while continually reminding me to concentrate on what is essential.

Abstract

Data-driven maintenance bears the potential to realize various benefits based on multifaceted data assets generated in increasingly digitized industrial environments. By taking advantage of modern methods and technologies from the field of data science and analytics (DSA), it is possible, for example, to gain a better understanding of complex technical processes and to anticipate impending machine faults and failures at an early stage. However, successful implementation of DSA projects requires multidisciplinary expertise, which can rarely be covered by individual employees or single units within an organization. This expertise covers, for example, a solid understanding of the domain, analytical method and modeling skills, experience in dealing with different source systems and data structures, and the ability to transfer suitable solution approaches into information systems. Against this background, various approaches have emerged in recent years to make the implementation of DSA projects more accessible to broader user groups. These include structured procedure models, systematization and modeling frameworks, domain-specific benchmark studies to illustrate best practices, standardized DSA software solutions, and intelligent assistance systems.

The present thesis ties in with previous efforts and provides further contributions for their continuation. More specifically, it aims to create supportive artifacts for the selection, evaluation, and application of data-driven methods in the field of industrial maintenance. For this purpose, the thesis covers four artifacts, which were developed in several publications. These artifacts include (i) a comprehensive systematization framework for the description of central properties of recurring data analysis problems in the field of industrial maintenance, (ii) a text-based assistance system that offers advice regarding the most suitable class of analysis methods based on natural language and domain-specific problem descriptions, (iii) a taxonomic evaluation framework for the systematic assessment of data-driven methods under varying conditions, and (iv) a novel solution approach for the development of prognostic decision models in cases of missing label information.

Individual research objectives guide the construction of the artifacts as part of a systematic research design. The findings are presented in a structured manner by summarizing the results of the corresponding publications. Moreover, the connections between the developed artifacts as well as related work are discussed. Subsequently, a critical reflection is offered concerning the generalization and transferability of the achieved results. Thus, the thesis not only provides a contribution based on the proposed artifacts; it also paves the way for future opportunities, for which a detailed research agenda is outlined.

Zusammenfassung

Datengetriebene Instandhaltung birgt das Potential, aus den in Industrieumgebungen vielfältig anfallenden Datensammlungen unterschiedliche Nutzeneffekte zu erzielen. Unter Verwendung von modernen Methoden und Technologien aus dem Bereich Data Science und Analytics (DSA) ist es beispielsweise möglich, das Verhalten komplexer technischer Prozesse besser nachzuvollziehen oder bevorstehende Maschinenausfälle und Fehler frühzeitig zu erkennen. Eine erfolgreiche Umsetzung von DSA-Projekten erfordert jedoch multidisziplinäres Expertenwissen, welches sich nur selten von einzelnen Personen bzw. Einheiten innerhalb einer Organisation abdecken lässt. Dies umfasst beispielsweise ein fundiertes Domänenverständnis, Kenntnisse über zahlreiche Analysemethoden, Erfahrungen im Umgang mit verschiedenen Quellsystemen und Datenstrukturen sowie die Fähigkeit, geeignete Lösungsansätze in Informationssysteme zu überführen. Vor diesem Hintergrund haben sich in den letzten Jahren verschiedene Ansätze herausgebildet, um die Durchführung von DSA-Projekten für breitere Anwendergruppen zugänglich zu machen. Dazu gehören strukturierte Vorgehensmodelle, Systematisierungs- und Modellierungsframeworks, domänenspezifische Benchmark-Studien zur Veranschaulichung von Best Practices, Standardlösungen für DSA-Software und intelligente Assistenzsysteme.

An diese Arbeiten knüpft die vorliegende Dissertation an und liefert weitere Artefakte, um insbesondere die Selektion, Evaluation und Anwendung datengetriebener Methoden im Bereich der industriellen Instandhaltung zu unterstützen. Insgesamt erstreckt sich die Abhandlung auf vier Artefakte, die in einzelnen Publikationen erarbeitet wurden. Dies umfasst (i) ein umfangreiches Systematisierungsframework zur Beschreibung zentraler Ausprägungen wiederkehrender Datenanalyseprobleme im Bereich der industriellen Instandhaltung, (ii) ein textbasiertes Assistenzsystem, welches ausgehend von natürlichsprachlichen und domänenspezifischen Problembeschreibungen eine geeignete Klasse von Analysemethoden vorschlägt, (iii) ein taxonomisches Evaluationsframework zur systematischen Bewertung von datengetriebenen Methoden unter verschiedenen Rahmenbedingungen sowie (iv) einen neuartigen Lösungsansatz zur Entwicklung von prognostischen Entscheidungsmodellen im Fall von eingeschränkter Informationslage.

Die Konstruktion der Artefakte wird durch einzelne Forschungsziele im Rahmen eines systematischen Forschungsdesigns angeleitet. Neben der Darstellung der einzelnen Forschungsbeiträge unter Bezugnahme auf die erzielten Ergebnisse der dazugehörigen Publikationen werden auch die Verbindungen zwischen den entwickelten Artefakten beleuchtet und Zusammenhänge zu angrenzenden Arbeiten hergestellt. Zudem erfolgt eine kritische Reflektion der Ergebnisse hinsichtlich ihrer Verallgemeinerung und Übertragung auf andere Rahmenbedingungen. Dadurch liefert die vorliegende Abhandlung nicht nur einen Beitrag anhand der erzeugten Artefakte, sondern ebnet auch den Weg für fortführende Forschungsarbeiten, wofür eine detaillierte Forschungsagenda erarbeitet wird.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial intelligence
C-MAPSS	Commercial modular aero-propulsion system simulation
CBM	Condition-based maintenance
CNN	Convolutional neural network(s)
CRISP-DM	Cross-industry standard process for data mining
DMME	Data mining methodology for engineering applications
DSA	Data science & analytics
DSR	Design science research
DSS	Decision support system(s)
HI	Health index
IAS	Intelligent assistance system(s)
IS	Information systems
KDD	Knowledge discovery in databases
LOWESS	Locally weighted scatterplot smoothing
LSTM	Long short-term memory
MAE	Mean absolute error
ML	Machine learning
PdM	Predictive maintenance
RMSE	Root mean square error
RNN	Recurrent neural network(s)
RO	Research objective
RUL	Remaining useful life
TDF	Time-domain feature(s)
TFDF	Time-frequency domain feature(s)

1 Introduction

"Data are becoming the new raw material of business."

Craig Mundie, Senior Advisor to CEO at Microsoft.

1.1 Motivation

The rapid advancements in computing power, sensors, storage engines, and internet technologies have a massive impact on our society and revolutionize the way we live, act, and work together. Business environments become increasingly digitized, and the ubiquitous use of IT has become an indispensable anchor for many organizations (Bley et al. 2016; Fichman et al. 2014). This situation favors the collection of vast amounts of data that can be generated with high frequency from multiple sources and heterogeneous systems (Chen et al. 2012; Constantiou and Kallinikos 2015). Within an organizational context, this kind of ubiquitously generated data can be seen as a valuable asset to establish data-driven business processes and fact-based decision making (Abbasi et al. 2016; Zschech et al. 2017). Empirical value propositions of data utilization include, for example, higher transparency, improved performance measurement, and the support and replacement of human decision making with automated algorithms (Wamba et al. 2015).

To exploit such potential and turn data into value, methods and tools of modern data analysis are required that are often subsumed under the collective term *data science and analytics* (DSA). In this thesis, DSA is defined as an analytical approach combining expertise from multiple disciplines, such as information systems (IS), computer science, statistics, and corresponding application domains, in order to discover meaningful relationships and hidden patterns from heterogeneous, multi-sourced data that can be converted into actionable insights (Agarwal and Dhar 2014; Ayankoya et al. 2014; Ramannavar and Sidnal 2016). Closely related to this approach is the term *data-driven methods*. Hereinafter, this term is defined as any systematic procedure that serves the purpose of processing data in order to achieve a certain analytical goal. This may range from simple techniques for calculating and visualizing descriptive indicators to more advanced algorithms from the field of machine learning (ML) that can automatically identify non-linear and complex relationships in high-dimensional data collections (Stefani and Zschech 2018).

A promising area for the application of DSA is the manufacturing sector. A decade ago, Manyika et al. (2011) had already estimated an amount of about two exabytes of newly generated data for just a single year. This provides a fundamental basis for improving various areas of interest such as quality control, process performance, production scheduling, and industrial maintenance (Brodsky et al. 2015; Flath and Stein 2018; Manyika et al. 2011). The area of maintenance is of particular interest since today's industry is characterized by increasingly complex production systems and machinery that require sophisticated maintenance systems to guarantee low environmental risks, high reliability, and human safety. Simultaneously, it is crucial to employ system functionalities and methods that allow efficient use of given resources and avoid unnecessary expenditures (Elattar et al. 2016; Peng et al. 2010). For this concern, the amount and the variety of data is vital, ranging from condition monitoring data and machine configurations to transactional records and event logs reflecting process executions. Such multifaceted data provide an ideal starting point for improved decision support and the discovery of unknown potentials. For example, technical processes can be better understood during health assessment; anomalous signs of degradation can be traced back to their root causes, and faults and failures can be anticipated at an early stage (Accorsi et al. 2017; Manyika et al. 2011). Thus, it is reasonable to assume that DSA applications in industrial maintenance offer excellent opportunities to extract hidden knowledge and make better use of given resources.

However, applying DSA in industrial settings is not a trivial task. Especially for technology users who might have rich domain expertise but lack sufficient DSA qualification, there are several hurdles to overcome. Often, there is no “silver bullet” approach that addresses a particular decision support task with a universal solution. Instead, DSA projects are usually iterative and time-consuming endeavors, and profound knowledge is required (i) to identify a suitable set of data-driven methods, (ii) to assess the methods’ results in a comprehensive manner, and (iii) to implement the chosen methods in practical settings under real conditions.

Against this background, this thesis aims to contribute several artifacts that mainly support the steps of selecting, evaluating, and applying data-driven methods in the field of industrial maintenance for the overall purpose of better decision support. The intended target groups of these artifacts are both practitioners and researchers working in manufacturing-related domains who require support and guidance for diving into the field of data science and analytics.

1.2 Conceptual Background

The maintenance function plays a fundamental role in today’s industrial value creation. It is concerned with all technical and administrative activities necessary to keep physical assets in their desired operating condition and to conduct countermeasures in case of deviations (Muchiri et al. 2011). Closely related to this fundamental principle, a variety of objectives can be pursued. These objectives include, for example, to ensure a system’s reliability and high product quality, to minimize machine downtime and risk of failure or damage, and to preserve plant safety, environmental protection, and resource efficiency (Horn and Zschech 2019).

In order to adequately meet such superior objectives, the central decision-making task of maintenance is determining the appropriate time at which necessary maintenance actions should be carried out. If actions are performed too late, i.e., after a fault or failure has occurred (also known as *corrective maintenance*), the result may include environmental risks, safety issues, machinery breakdowns, and impaired product quality. If, by contrast, actions are carried out too early, for example, due to fixed periodic intervals (also known as *preventive* or *time-based maintenance*), high expenses may arise as a result of regular interventions or unused service lifetime (Peng et al. 2010; Veldman et al. 2011).

To address this crucial trade-off, a more proactive decision-making strategy has emerged, called *condition-based maintenance* (CBM). In this strategy, comprehensive data are gathered and processed by a condition monitoring system to assess the current state of the equipment and derive recommendations for the optimal time and type of intervention (Jardine et al. 2006). Figure 1 illustrates the relationship between maintenance costs, reliability, and the remaining useful life

(RUL) of a system, and it assigns all three maintenance strategies. Following the CBM approach, divergent machine behavior can be detected and classified at an early stage through diagnostic techniques in order to reduce the uncertainty of maintenance actions and avoid unnecessary work by taking actions only when there is evidence of anomalous behavior. Furthermore, by using suitable indicators and prognostic techniques, it is possible to determine the machine's future state or its RUL, which is often also referred to as *predictive maintenance* (PdM) (Elattar et al. 2016; Ran et al. 2019).

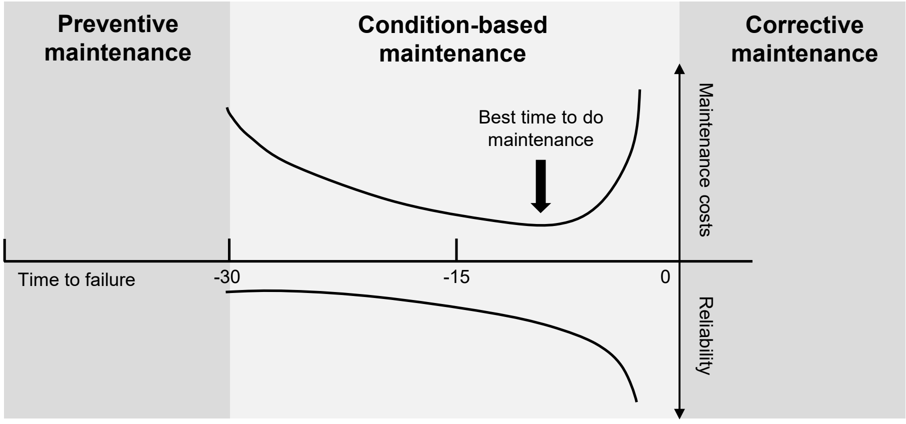


Figure 1: Relationship between maintenance costs, reliability, and RUL (adapted from Peng et al. 2010)

For the development of diagnostic and prognostic maintenance models, three basic types of approaches are applicable. The first type is composed of *physical model-based approaches*. Here, mathematical models of physical processes are developed by experts in the field, and large sets of data observations validate the parameters of the model. Such models generally have the advantage of being very accurate since they are based on natural laws (e.g., specific degradation laws). However, their development can be considered as costly and time-consuming because it requires a thorough understanding of the physical mechanisms of the system under consideration.

The second type includes *knowledge-based approaches* that try to simulate human thinking. A representative example is that of expert systems in which domain knowledge from human specialists is formalized in terms of rules in order to allow automated reasoning. While such systems provide a useful form of encapsulating human expertise, it is challenging to obtain such knowledge and convert it into adequate rules (Elattar et al. 2016; Peng et al. 2010).

The third type comprises purely *data-driven approaches*. In this category, extensive data collection is exploited using techniques from disciplines like statistics or ML in order to automatically extract patterns and relationships of interest. In contrast to physical models and knowledge-based approaches, data-driven methods have the advantages that (i) they do not require comprehensive system knowledge, (ii) they are relatively fast to implement, (iii) they can be tuned for similar systems, and (iv) they can exploit hidden relations and nuances within the data records (Elattar et al. 2016; Peng et al. 2010).

Due to the given advantages, data-driven methods have proven to be a promising alternative when implementing maintenance decision models, which probably is why they are gaining increasing attention in research and industry (Ran et al. 2019). Nevertheless, several aspects are hampering the selection, evaluation, and application of data-driven methods in practical settings. Some of the critical factors, as observed by the author of this thesis, are summarized below.

Heterogeneity of decision support tasks: Although the maintenance function at its core is "only" concerned with the central task of determining the appropriate time for intervention, there are several different facets related to this task. For example, both diagnostic and predictive issues can be further broken down into several sub-aspects, which in turn require different approaches and methods for their implementation (Jardine et al. 2006). Additionally, there are descriptive as well as prescriptive analytic tasks that further complement the field and thus increase heterogeneity (Karim et al. 2016).

Heterogeneity of data-driven methods: The body of knowledge on data-driven maintenance is extensive, as already noted by Jardine et al. (2006). Hundreds of papers are published every year by researchers and developers from multiple scientific communities, such as computer science and engineering disciplines, bringing forth a variety of analytical methods for diverse contexts. Such methods range from statistical analysis and mathematical modeling to algorithms from ML and data mining (DM). Each of these allows access to a diversity of data from multiple perspectives with individual merits and limitations (Accorsi et al. 2017; Ran et al. 2019).

Heterogeneity of maintenance-related data: The quantity and variety of data have increased considerably due to (i) the growing complexity of machinery consisting of multiple components, (ii) the ubiquitous embedding of modern sensor technology, and (iii) the linkage with various adjacent application systems (Manyika et al. 2011; Ran et al. 2019). Past decision support was mainly based on time-series signals from sensors such as pressure, vibration, and temperature (Jardine et al. 2006). In the meantime, however, more data types can be gathered and processed, such as demonstrated by Sipos et al. (2014) and Gutschi et al. (2019), where the authors used event logs (e.g., process executions, textual error messages) for machine failure prediction.

Limited availability of representative data: Despite the high availability of multifaceted data collection, there is often crucial information lacking in industrial settings that is necessary for adequate decision support. Especially, supervised ML methods that learn relationships from many historical observations require representative training data reflecting a system's characteristic behavior from normal and faulty operations to degradation patterns under multiple operating conditions. Such "run-to-failure" data are often scarce in industry and can only be procured at great expense due to zero-downtime policies (Leturiondo et al. 2017; Susto et al. 2015).

Multidisciplinary skill requirements: For the implementation of data-driven methods, multidisciplinary DSA skills covering a joint consideration of the aspects mentioned above are required. Hence, this includes a solid understanding of the domain, expertise with numerous analytical methods, experience with different data sources, and the ability to transfer results into technological solutions based on advanced programming and software engineering skills (Schumann et al. 2016; Zschech et al. 2018). However, fully equipped DSA professionals covering all these requirements are still a rare species, whereas conducting data-driven projects in interdisciplinary teams with multiple experts remains an iterative and time-consuming endeavor (Hesenius et al. 2019; Huber et al. 2019).

1.3 Related Work

In order to address the challenges above and provide support for the selection, evaluation, and application of data-driven methods in general, and in industrial maintenance in particular, several research efforts have been undertaken in recent years, bringing forth a variety of contributions related to this thesis. In the following, some of these efforts are briefly outlined.

Surveys and systematizations: Given the plethora of research on data-driven maintenance, there is also a high number of literature surveys. They help to structure the field and provide systematizations to classify the broad number of methods from different perspectives. For example, Bousdekis et al. (2018) identified several methods for predictive and prescriptive tasks. They organized them into a structured framework to guide the selection of suitable method combinations by considering the desired output, the given (data) input, and the availability of domain knowledge. Other surveys assess the applicability requirements of reviewed methods (e.g., Javed et al. 2017) or describe their merits and limitations (e.g., Heng et al. 2009; Ran et al. 2019) in order to guide the selection of suitable methods in practical settings.

Models for recurring data analysis problems: Similar to the method selection framework by Bousdekis et al. (2018), there are a few more attempts to describe and model recurring problem classes for which generic solution templates can be applied. For example, Brodsky et al. (2015) developed a software framework for DSA solutions based on a reusable knowledge base for solving recurring analytical tasks in production environments. Similarly, Eckert and Ehmke (2017) propose the standardization of data analysis tasks in industrial settings by constructing a reference model. On a more general level, Russo (2016) introduces the vision of so-called “data analysis patterns” as an analogy to design patterns in software engineering. Such patterns could be considered as guiding models or templates to instruct users on how to apply an intentional solution design for recurring data analysis problems based on accumulated experiences instead of rediscovering a problem solution every time from scratch. However, little research has been done in this particular context so far. Some inspiring exceptions are the research efforts by Nalchigar et al. (e.g., Nalchigar et al. 2019; Nalchigar and Yu 2020). The authors propose a comprehensive conceptual modeling framework for DSA solution patterns, which among other elements consists of different modeling views (i.e., business questions, analytics design, data preparation), view-specific design catalogs, a metamodel, and several application examples.

Structured procedure models: Procedure models organize tasks and activities of design and development processes into structured, logically arranged steps in which corresponding methods and techniques are applied. In the area of DSA and DM, several such procedure models have been developed to provide instructions for all relevant phases from domain and data understanding to data preparation, analytical method selection, and evaluation (Mariscal et al. 2010). Prominent examples are the CRISP-DM methodology (cross-industry standard process for data mining) (Wirth and Hipp 2000) and the KDD (knowledge discovery in databases) process model (Fayyad et al. 1996). While such models offer generic guidance across different branches, their applicability in concrete cases often suffers from a lack of domain specificity. For this purpose, Huber et al. (2019) proposed an extended CRISP-DM version particularly tailored for production domains. The authors integrate additional phases for technical understanding, realization, and implementation in order to address specific application scenarios such as predictive maintenance and process optimization.

Standardized DSA software and intelligent assistance systems: In order to make data analysis projects more accessible to broader user groups, especially to DSA novices with little background knowledge, a variety of software solutions have emerged over the past fifty years such as SAS, SPSS, and KNIME. They offer standardized functionalities and therefore require no programming skills. Simultaneously, a high diversity of intelligent assistance systems (IAS) have evolved that are increasingly integrated into standardized DSA software. Such IAS offer different kinds of features to guide users through all stages of the data analysis process and simplify the selection, evaluation, and application of analysis operators and their results (Serban et al. 2013). Two illustrative examples of such features can be observed in the DSA platform RapidMiner, called "Auto Model" and "Wisdom of Crowds" (RapidMiner 2020). The first feature takes a dataset as input and then automatically suggests the best performing ML technique for a particular task. The second example is built upon a best-practice knowledge base derived from the activities of more than 250,000 platform users to recommend suitable analysis operators and parameters within a data analysis workflow.

Public benchmark datasets: Due to the scarce availability of run-to-failure data that constitute a prerequisite for many data-driven prognostic methods, there have been several initiatives to generate synthetic datasets for research and education purposes. Prominent examples come from NASA's Prognostics Data Repository[1] and include datasets from different technical settings such as milling machines, bearings, turbofan engines, and battery charging cycles (Eker et al. 2012; Lei et al. 2018). Derived from laboratory experiments and advanced simulations, such synthetic datasets usually show realistic properties. Therefore, they provide a fundamental basis for the development and assessment of data-driven prognostic solutions. Accordingly, they are frequently used by researchers and practitioners as objective benchmark settings and for teaching purposes to demonstrate the merits and limitations when comparing different methods.

1.4 Research Design

Inspired by the potentials of DSA applications in maintenance scenarios and the previous efforts in related work, this thesis aims to provide additional contributions to complement the field and further promote the use of data-driven methods in industrial environments. For this purpose, a systematic research design is proposed that is concerned with the overall research objective, *to create supportive artifacts for the selection, evaluation, and application of data-driven methods in the field of industrial maintenance*. As illustrated in Figure 2, the overarching objective is expressed through four more specific research objectives (**RO1-RO4**) that relate to certain focus areas. Moreover, they are further refined by individual sub-objectives. In order to achieve these objectives, well-established research methods are applied, which in turn are refined by individual method components.

---

[1] https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/ (last access: 01-06-2020)

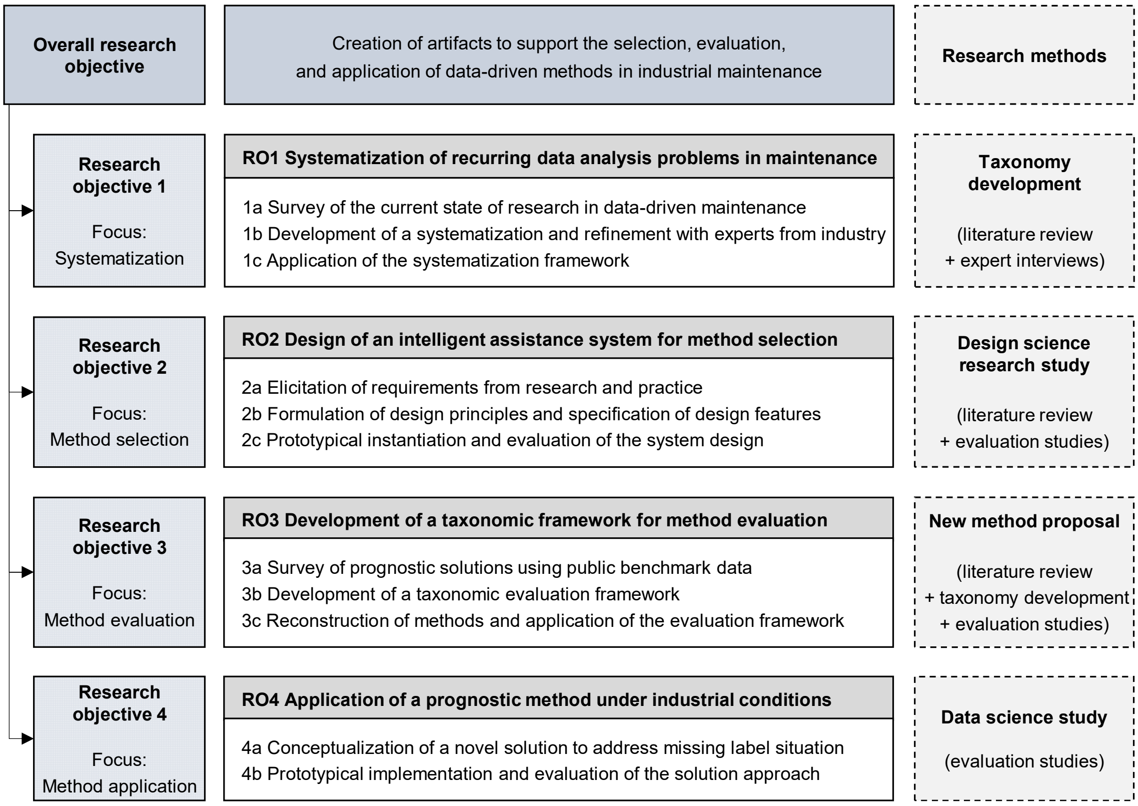


Figure 2: Summary of the research design

The first research objective (**RO1**) aims at a systematization of the field with a particular focus on the dimensions and characteristics of recurring data analysis problems in maintenance settings. More specifically, it is intended to take into account the findings of related work and extract descriptive elements to classify the broad variety of data analysis objectives, data assets, and analytical methods (1a). Based on the findings, a structured systematization framework shall be developed, which will be further refined by expert knowledge from industry to additionally reflect the practitioners' points of view (1b). An exemplary application of the final systematization framework shall subsequently demonstrate the usefulness of the created artifact (1c). In order to conduct this type of research, a taxonomy development approach is pursued (Nickerson et al. 2013), in which a systematic literature review (vom Brocke et al. 2009) and semi-structured expert interviews (Myers and Newman 2007) are embedded.

The second research objective (**RO2**) is concerned with the particular aspect of method selection. While there have been several efforts for guiding the task of method selection, especially in an automated manner using different types of IAS, only a few approaches take into account the particularities of a problem context expressed in a domain-specific language to select a suitable method. Such an approach could help domain experts stay in their familiar surroundings without the need to acquire more profound DSA knowledge when starting to implement data-driven projects (Eckert and Ehmke 2017; Hogl 2003). For this purpose, the design of a novel IAS shall be proposed that takes problem descriptions articulated in natural language as input and offers advice regarding the most suitable class of DM methods to address the problem. Following a design science research (DSR) methodology for this approach (Peffers et al. 2007), the research objective is further divided into three parts: (i) the elicitation of requirements from research and practice (2a), (ii) the proposal of design principles and specified design features (2b), and (iii) the evaluation of the system design based on a prototypical instantiation (2c).

The third research objective (**RO3**) addresses the development of a new evaluation framework to assess data-driven methods and solutions more systematically and comprehensively. Especially in the area of prognostic solution development, where a large proportion of studies are based on public benchmark datasets, most solutions or method pipelines are evaluated using only a single score to assess whether they perform better or worse than existing approaches. While a single score proves to be the right choice for a quick and aggregated comparison, there is a lack of transparency about which particular components, such as specific pre-processing and modeling steps, affect the overall performance. Thus, inspired by the methodical taxonomy approach applied for RO2 to create a systematization, the potential could be discovered to modify and re-apply this approach for the decomposition of data-driven solutions into taxonomic components. This helps to reduce their complexity and allows an evaluation on a more fine-grained basis. Accordingly, a new method proposal is offered that consists of (i) a literature survey procedure to identify prognostic solutions based on public benchmark data (3a), (ii) a refined taxonomy development approach to create a framework with modular components of data-driven solutions (3b), and (iii) quantitative evaluation studies to reconstruct the identified prognostic solutions and apply the framework for a more fine-grained method evaluation (3c).

The fourth research objective (**RO4**) deals with the aspect of method application in real-world environments. In contrast to synthetical settings, real production environments often lack representative training data for the establishment of prognostic decision models. In the case of critical machines, for example, the aim is to avoid failures through strictly short maintenance intervals. Moreover, it is often not possible to carry out test runs that go beyond the limits of safe conditions due to the pressure to use plants efficiently (Leturiondo et al. 2017; Susto et al. 2015). This situation results in “missing labels”, which can be seen as a significant hurdle in the development of adequate prognostic models (Gouriveau et al. 2013). To address this problem and show how it is possible to provide maintenance decision support in this unfortunate situation, a novel solution approach shall be developed. For this purpose, a real-world case of a German car manufacturer facing an imperfect maintenance situation is taken as an example to conduct a data science study for solution development (Mariscal et al. 2010). The challenge of the case is to support the decision-making process of a wear-induced tool replacement in a milling machine by predicting the tools’ RUL when no labels are present due to individual risk preferences and poor information available. To this end, the fourth research objective is structured into two parts. The first part includes the conceptualization of a novel solution (4a). The second part covers the prototypical implementation and an evaluation to assess the approach’s feasibility (4b).

1.5 Structure of the Thesis

In order to address the proposed research objectives, the remaining thesis is organized into four main chapters and two additional chapters. Each main chapter is represented by individual publications written and published from 2017 to 2020 as part of an accumulative research process. The internal structure of the main chapters follows the composition of the sub-objectives from the research design. Thus, after a short recapitulation of the topic and relevant background information, the achieved results of the publications are summarized for each research objective. Moreover, in some sub-chapters, further elaborations are outlined, which have not yet been subject to published work (e.g., the exemplary application of the frameworks).

Figure 3 summarizes the structure of the thesis and indicates how each publication contributes to the achievement of the research objectives. **Chapter 2** covers the development of a framework to systematize the field (RO1) based on the findings of publication P1 (cf. Appendix II: A). **Chapter 3** is concerned with the design of the new IAS for automated method selection (RO2) as a result of publications P2 and P3 (cf. Appendix II: B and C). **Chapter 4** focuses on the creation of a novel evaluation framework (RO3) by referring to publication P4 (cf. Appendix II: D), and **Chapter 5** addresses the topic of prognostic method application under industrial conditions (RO4) by reflecting the results of publication P5 (cf. Appendix II: E).

After the four main chapters, **Chapter 6** offers a discussion of the results. First, this includes a consideration of connections between the individual artifacts and related work. Subsequently, the results are critically reflected concerning their generalization and transferability in order to highlight achieved contributions as well as prospects for future work. Finally, some concluding remarks are provided in **Chapter 7**.

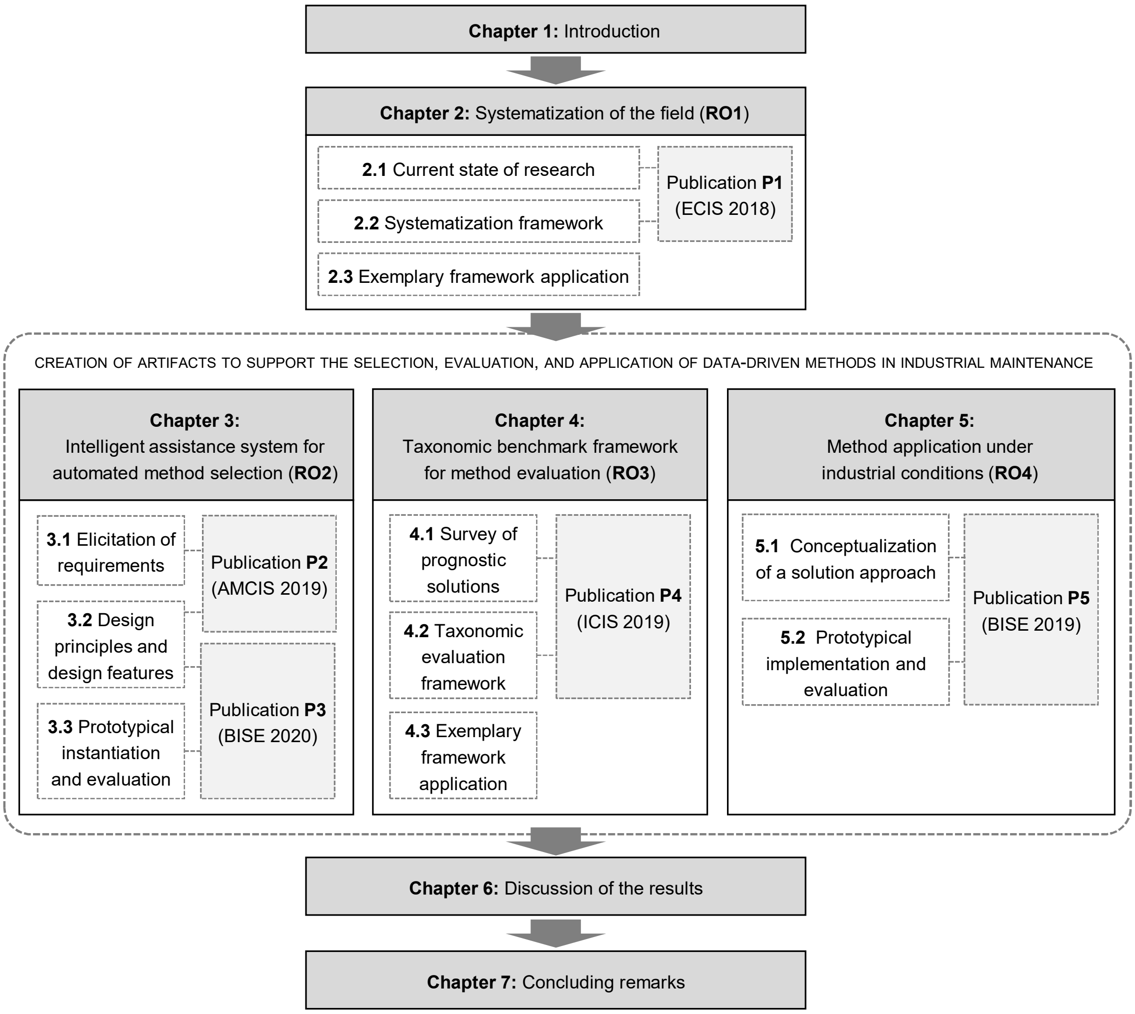


Figure 3: Structure of the thesis

2 Systematization of the Field

Research objectives	
RO1	Systematization of recurring data analysis problems in maintenance
1a	Survey of the current state of research in data-driven maintenance
1b	Development of a systematization and refinement with experts from industry
1c	Application of the systematization framework
Reference to original work	
Zschech (2018)	Publication P1	Appendix II: A

Table 1: Research summary for Chapter 2

This chapter addresses the first research objective: to create a systematization that organizes dimensions and characteristics of recurring data analysis problems in data-driven maintenance scenarios within a structured framework (cf. Table 1). For this purpose, a taxonomy development approach was chosen. In general, taxonomies serve as viable tools for organizing knowledge in a structured manner and manifesting descriptive theories (Gregor 2006). As such, they enable researchers to study the relationship among concepts and help to analyze and understand complex domains (Nickerson et al. 2013).

To carry out the taxonomy development, the research method proposed by Nickerson et al. (2013) was applied as it provides systematic guidance. It basically consists of three steps: (i) determining a meta-characteristic, (ii) specifying ending conditions, and (iii) identifying dimensions and characteristics of the taxonomy. The meta-characteristic is the root element and serves as the foundation for the choice of all the other characteristics. For this purpose, a tripartite meta-characteristic was chosen to distinguish between *data analysis objectives* (describing the output of a problem to be solved), *data characteristics* (describing the input), and *analytical techniques* (describing the actual steps of data processing to achieve the objectives) (Tsai et al. 2014). Subsequently, the specification of ending conditions was required due to the iterative method. To this end, a variety of criteria can be defined in order to fulfill specific quality properties such as robustness and conciseness of the taxonomy (Nickerson et al. 2013).

The actual step of identifying dimensions and characteristics can then be carried out either with an empirical-to-conceptual or a conceptual-to-empirical path, where it is recommended to combine both paths for the integration of different perspectives. Accordingly, this procedure was organized in multiple iterations. Conceptual knowledge was derived from the vast body of research in the academic literature, while empirical knowledge was collected through interviews with DSA experts from industry. In the following, the results of the taxonomy development are briefly described following the structure of the sub-objectives. Thus, it starts with a reflection of the conducted literature review (1a). Then, the taxonomy structure is described using additional findings from the expert interviews (1b). Finally, the retrieved framework is applied to two distinct cases in order to demonstrate the usefulness of the created framework (1c).

2.1 The Current State of Research

To examine the current state of research in data-driven maintenance, a systematic literature review was conducted using multiple digital libraries such as ScienceDirect, ACM, and IEEE Xplore (vom Brocke et al. 2009). More specifically, two review cycles were carried out to retrieve relevant literature. The first cycle was limited to a search for papers in connection with the concept of "maintenance analytics" in order to identify studies that propose similar frameworks for the systematization of data analysis problems from a DSA perspective. However, only a few papers with a limited scope could be identified (e.g., Famurewa et al. 2017; Karim et al. 2016).

In the second review cycle, the search procedure considered concepts related to CBM and PdM, which generally show a broader coverage in the scientific community of data-driven maintenance. In this way, a large body of knowledge could be studied. For example, searching just the digital library of ScienceDirect yielded more than 5,000 results (3,063 hits for PdM and 2,103 hits for CBM, day of search: 08-08-2017). The results covered different types of literature, ranging from context-specific solutions to conceptual discussions of CBM and PdM programs. Moreover, the search results included a large number of survey papers, which were of particular importance as they summarize the field from multiple perspectives. By taking the results from all digital libraries together, a total of 99 survey papers were found with an emphasis on different maintenance technologies, models, and algorithms for data processing and decision making. After reviewing all the papers, the number of relevant items was reduced by 79, as most of the surveys deal with specific aspects such as (i) particular application domains (e.g., railway or wind turbines), (ii) specific machine components (e.g., power transformers), and (iii) other individual aspects (e.g., cloud-based approaches). The remaining 20 articles[2], on the other hand, offer a broad and comprehensive summary of the field, including various systematizations of how decision tasks, data-driven methods, and data inputs can be classified. However, a more detailed analysis revealed a highly diffuse picture, especially as to classifying the extensive amount of available methods. For this reason, it was necessary to harmonize the existing systematizations to some extent in order to obtain a structured taxonomy framework.

2.2 Systematization Framework

By using the identified literature from both review cycles, it was possible to iteratively create a first taxonomy draft to distinguish between numerous dimensions and characteristics of data analysis problems in maintenance. For example, data analysis objectives could be divided into four distinct types (i.e., descriptive, diagnostic, prognostic, and prescriptive), each of them consisting of further sub-types. Data assets could be basically grouped into event data and condition-monitoring data, from which further properties could be derived (e.g., monitoring frequency, event types). For analytical methods and techniques, on the other hand, a more

---

[2] Ahmad and Kamaruddin (2012); Ahmadzadeh and Lundberg (2014); An et al. 2015; Ao (2011); Bousdekis et al. (2018); Dragomir et al. (2009); Elattar et al. (2016); Goyal and Pabla (2015); Hashemian and Bean (2011); Heng et al. (2009); Jardine et al. (2006); Kothamasu et al. (2006); Lee et al. (2014); Peng et al. (2010); Prajapati et al. (2012); Schwabacher (2005); Si et al. (2011); Veldman et al. (2011); Vogl et al. (2019); Zhu et al. (2016)

versatile picture was observed. This situation was also noted by Elattar et al. (2016) when reviewing typologies for prognostic methods: "Sometimes, the classification is based on the type of available data and knowledge about the system. Another time prognostics approaches are classified according to the type of the used methodology" (p. 132). For the latter part, for example, some authors generally distinguish between statistical models and artificial intelligence (AI) (e.g., An et al. 2015; Peng et al. 2010). Other authors use a more fine-grained differentiation and group methods, for example, into regression-based methods, trend projection methods, reliability-based methods, and filtering-based methods (e.g., Heng et al. 2009; Si et al. 2011).

As a result of merging the heterogeneous classification schemes identified throughout the surveys, more than 80 characteristics were identified that could be grouped and organized within more than 25 dimensions. Consequently, the taxonomy lacked being sufficiently concise and comprehensive because it consisted of too many unstructured, partially overlapping dimensions.

For this reason, multiple interviews with experts from industry were carried out to additionally consider empirical knowledge from practitioners and see how real-world scenarios could be mapped onto the taxonomy draft. In particular, seven DSA professionals working for a medium-sized IT service provider were recruited. To conduct the interviews, a qualitative, semi-structured approach was applied (Myers and Newman 2007), addressing the following three aspects: (i) introduction to the research project, (ii) identification of contextual information and recurring properties of the interviewees' DSA projects in maintenance, and (iii) discussion and modification of the proposed taxonomy draft.

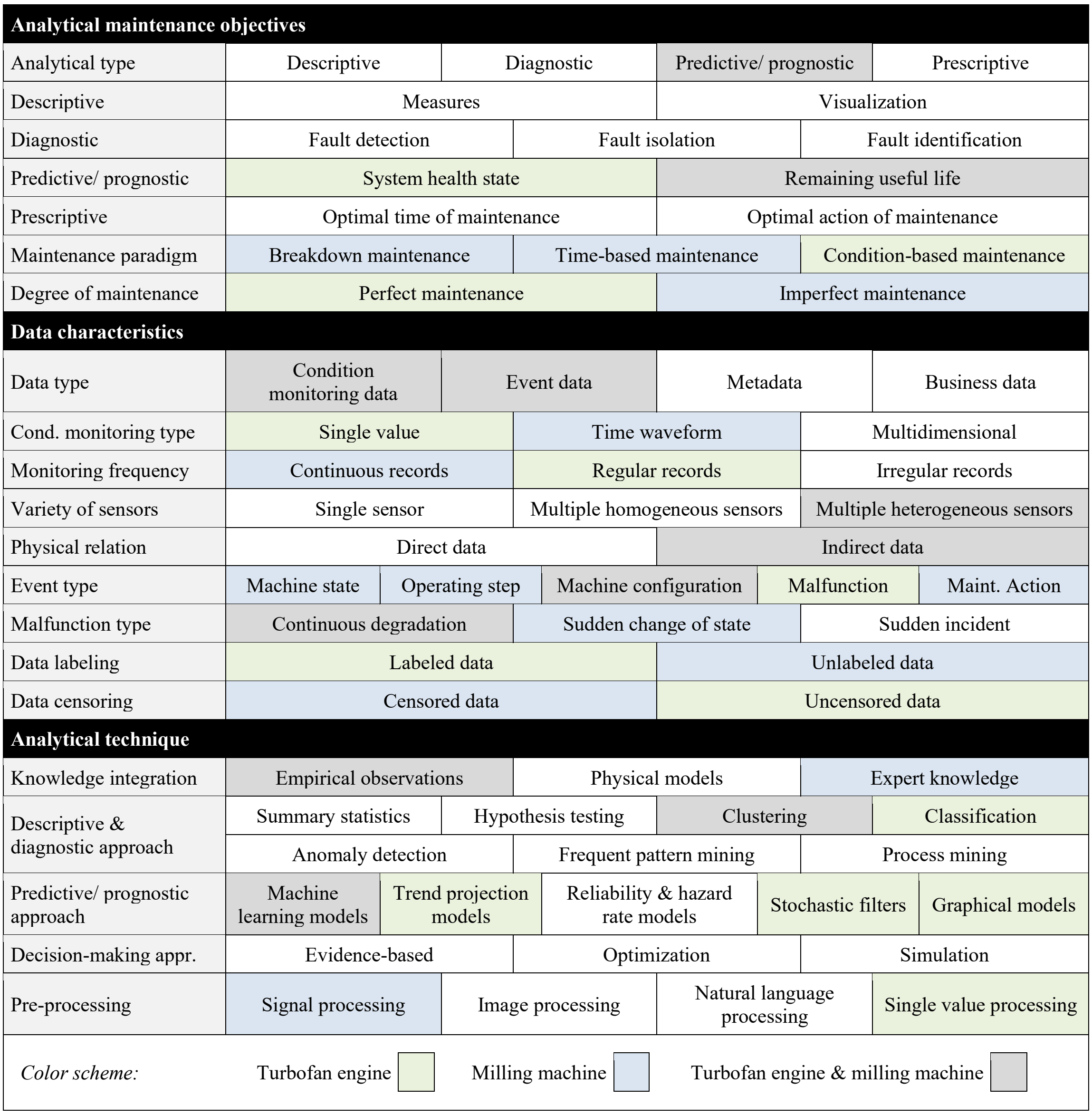
In this way, the systematization framework could be evaluated and enriched with experiences from industrial practice. Furthermore, it was possible to reduce the degree of complexity to make the taxonomy more precise. As a result, the final taxonomy covered 67 characteristics organized into 21 dimensions. The results are visualized in the next section. For a detailed description of the dimensions and characteristics, please refer to the full study (Zschech 2018).

2.3 Exemplary Framework Application

With the resulting taxonomy, it is possible to classify data analysis problems by their core characteristics in order to identify both commonalities as well as differences between different maintenance scenarios. For demonstration purposes, the framework application is illustrated below using two example cases, which are also the subject of later studies in this thesis.

The first case refers to a turbofan engine degradation scenario based on NASA's C-MAPSS data (commercial modular aero-propulsion system simulation) (cf. Chapter 4). This scenario is a commonly applied benchmark setting for which a simulation environment was used to generate synthetic datasets. Those datasets are made publicly available for the development of new prognostic solutions (Ramasso and Saxena 2014). The second case refers to the real-world setting of a German car manufacturer, where the step of replacing wear-induced tools in a milling machine should be supported through a proactive solution (cf. Chapter 5). More specifically, the aim was to predict the milling tools' RUL, with little information available due to missing quality thresholds and individual risk preferences of the machine operators.

For both cases, a classification of central characteristics is carried out using the derived systematization framework. Table 2 summarizes the results. The green shades indicate the turbofan engine degradation scenario, while the blue shades represent the scenario of the milling machine. Shared characteristics between the two cases are highlighted in grey.



Analytical maintenance objectives					
Analytical type	Descriptive	Diagnostic	Predictive/ prognostic	Prescriptive	
Descriptive	Measures	Visualization			
Diagnostic	Fault detection	Fault isolation	Fault identification		
Predictive/ prognostic	System health state	Remaining useful life			
Prescriptive	Optimal time of maintenance	Optimal action of maintenance			
Maintenance paradigm	Breakdown maintenance	Time-based maintenance	Condition-based maintenance		
Degree of maintenance	Perfect maintenance	Imperfect maintenance			
Data characteristics					
Data type	Condition monitoring data	Event data	Metadata	Business data	
Cond. monitoring type	Single value	Time waveform	Multidimensional		
Monitoring frequency	Continuous records	Regular records	Irregular records		
Variety of sensors	Single sensor	Multiple homogeneous sensors	Multiple heterogeneous sensors		
Physical relation	Direct data	Indirect data			
Event type	Machine state	Operating step	Machine configuration	Malfunction	Maint. Action
Malfunction type	Continuous degradation	Sudden change of state	Sudden incident		
Data labeling	Labeled data	Unlabeled data			
Data censoring	Censored data	Uncensored data			
Analytical technique					
Knowledge integration	Empirical observations	Physical models	Expert knowledge		
Descriptive & diagnostic approach	Summary statistics	Hypothesis testing	Clustering	Classification	
	Anomaly detection	Frequent pattern mining	Process mining		
Predictive/ prognostic approach	Machine learning models	Trend projection models	Reliability & hazard rate models	Stochastic filters	Graphical models
Decision-making appr.	Evidence-based	Optimization	Simulation		
Pre-processing	Signal processing	Image processing	Natural language processing	Single value processing	
Color scheme:	Turbofan engine 🟩	Milling machine 🟦	Turbofan engine & milling machine ⬜		

Table 2: Application of the systematization framework using two example cases

Concerning the characterization of analytical maintenance objectives, the main focus in both scenarios is to establish a prognostic decision model. More specifically, the central predictive task is concerned with RUL estimation, while some C-MAPSS studies have used the turbofan scenario for health state estimation (Ramasso and Saxena 2014). Regarding the observed maintenance paradigm and the degree of maintenance, the two cases differ. The milling case is an imperfect scenario with smaller corrections made until the milling tools are finally replaced. Due to missing condition monitoring thresholds, the tool replacements are performed either too late (i.e., similar to “breakdown” paradigm) or at regular intervals (i.e., time-based paradigm) (Zschech, Heinrich, Bink, et al. 2019).

By contrast, the turbofan scenario simulates an ideal CBM setting with run-to-failure data where the end of each cycle represents a critical threshold for RUL estimation. Moreover, the effects of between-flight interventions can be neglected as they are already incorporated within the simulated sensor measurements in the form of process noise. For this reason, the case can be classified as a perfect maintenance scenario (Saxena, Goebel, et al. 2008).

From a data perspective, the two cases have in common that they are characterized by event data about machine configurations as well as an extensive collection of indirect condition monitoring data from multiple heterogeneous sensors reflecting a continuous degradation. However, there are also several attributes distinguishing the two cases. More specifically, the C-MAPSS collection encompasses five datasets replicating the degradation behavior of turbofan engines under a variety of operating conditions and fault modes. Each dataset covers multiple turbofan engines and contains single-value snapshots of 21 sensor measurements for each simulated flight (i.e., cycles). It is assumed that each measurement is captured in regular intervals, i.e., either during or right after a flight. Varying operating conditions are a result of different machine configurations represented by three parameters that are individually specified for each flight. Furthermore, the last cycle of each engine can be considered as a “malfunction” event marking the end of useful life. Therefore, the datasets only contain uncensored run-to-failure samples with full label information for training purposes (Saxena, Goebel, et al. 2008).

By contrast, the milling scenario is subject to missing label information since no malfunction events can be obtained, which partly results from censored data records. These circumstances also constitute the core challenge to the case for the development of a prognostic model. Nevertheless, the case offers broad availability of other event data that can be used for solution development. These include event records about (i) machine states (e.g., running, finished) in order to derive information about produced units, (ii) operating steps (e.g., milling, cleaning) in order to focus on relevant phases, (iii) maintenance actions in order to distinguish between perfect and imperfect interventions, and (iv) machine configurations in order to track the changes made by parameter corrections. Another vital source of information is provided by the condition monitoring data, which, however, also differs from the turbofan engine case. Thus, sensor measurements of multiple milling components (e.g., machine axes and spindles) are continuously recorded for each operating step, resulting in fine-grained time waveform data. These measurements not only reveal a continuous degradation behavior of the milling tools; they also indicate sudden changes in machine conditions during material processing (Zschech, Heinrich, Bink, et al. 2019).

With the last meta-characteristic of the framework, it is possible to characterize different analytical techniques and methods applied for solution development. In the turbofan engine scenario, only empirical observations are used without additional sources of knowledge. The prognostic models of existing C-MAPSS studies are based on a variety of approaches such as ML, trend projection, stochastic filters, and graphical models. Additionally, in some studies, clustering and classification approaches are used as preparatory steps for health state prediction. Moreover, given the nature of the sensor measurements, single-value processing techniques are required for pre-processing (Ramasso and Saxena 2014).

In the milling scenario, on the other hand, it was possible to derive additional expert knowledge to confirm several preliminary findings during exploratory analysis steps. For solution development, different clustering approaches are used to discover hidden structures and extract useful label information, while ML models are applied for RUL estimation. Moreover, signal processing techniques are used to reduce the dimensionality of time waveform data given by the fine-grained sensor measurements (Zschech, Heinrich, Bink, et al. 2019).

In summary, it can be seen that the application of the systematization framework provides a quick overview to highlight central commonalities as well as distinctive properties between data analysis problems in different maintenance scenarios. This overview can guide various stakeholders involved in maintenance-related DSA projects. For example, modeling experts and data analysts can gain insights into the particularities of the domain represented by the data characteristics and the maintenance objectives. Domain experts, on the other hand, can better understand the analytical toolset for the technical implementation, referring to standards and best practices. In this way, the systematization framework serves as a viable instrument for communication purposes and for bringing together different actors to discuss a multidisciplinary problem space collectively.

3 Intelligent Assistance System for Automated Method Selection

Research objectives			
RO2	Design of an intelligent assistance system for automated method selection		
2a	Elicitation of requirements from research and practice		
2b	Formulation of design principles and specification of design features		
2c	Prototypical instantiation and evaluation of the system design		
Reference to original work			
	Zschech, Heinrich, Horn, et al. (2019)	Publication P2	Appendix II: B
	Zschech et al. (2020)	Publication P3	Appendix II: C

Table 3: Research summary for Chapter 3

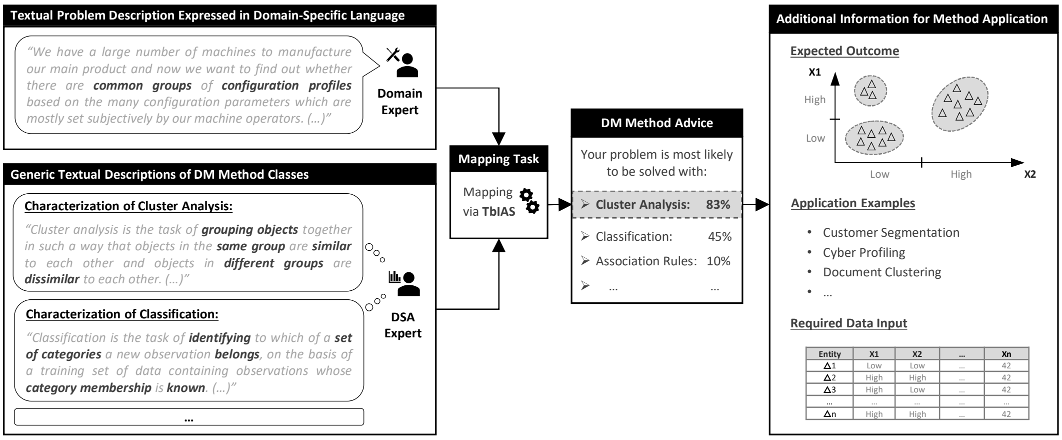
In any DSA project, the task of mapping a domain-specific problem onto an adequate set of DM methods by experts in the field is a crucial step. However, these experts may not always be available, and DM novices have to perform the task themselves. For this reason, there have been several research efforts towards automated method selection as a means of support. Most approaches are part of modern IAS (Serban et al. 2013) and can be roughly divided into three categories: (i) expert systems (e.g., Dabab et al. 2018; Danubianu 2008), (ii) meta-learning systems (e.g., Kerschke et al. 2019; Lemke et al. 2015), and (iii) question answering systems (e.g., Hogl 2003). However, none of the existing approaches operates on a suitable level of abstraction, and none can consider the particularities of problems expressed in the natural and domain-specific language of the novice. Therefore, this chapter is concerned with the second research objective, to propose the design of a novel IAS that takes problem descriptions articulated in natural language as input and offers advice regarding the most suitable class of DM methods.

In order to conduct this kind of research, a DSR approach was pursued. Design science is a fundamental paradigm in IS research as it is concerned with the construction of socio-technical artifacts to solve organizational problems and derive prescriptive design knowledge (Gregor and Hevner 2013). More specifically, the DSR procedure model proposed by Peffers et al. (2007) was adopted, consisting of six steps: (i) problem identification and motivation, (ii) definition of the objectives for a solution, (iii) design and development, (iv) demonstration, (v) evaluation, and (vi) communication. Please note, while publication P2 primarily focused on the first two steps and the preliminaries for step (iii), publication P3 covers the full DSR procedure and refines some of the previous results based on more recent findings. Therefore, the publications differ slightly concerning the adoption of the six steps.

In the following, the results of the general DSR approach are briefly described in order to achieve the defined sub-objectives 2a–2c (cf. Table 3). It starts with the elicitation of requirements from research and practice, followed by the design proposal in terms of design principles and design features. Finally, the system design is evaluated based on a prototypical instantiation.

3.1 Elicitation of Requirements

To obtain initial requirements, the mapping problem was conceptualized with a typical scenario, as observed in practice. In this scenario, a domain expert provides a problem description in natural language, and a DSA expert is consulted to realize a mapping with a specific class of DM methods using his or her knowledge about different methods. To support this task in an automated manner, a novel IAS should offer the functionality outlined in Figure 4. It receives a problem description and recognizes all relevant entities of interest. Built upon an advanced learning base, the IAS can then infer which class of DM methods most likely addresses the problem. On this basis, further information about the DM method is provided as guidance for its application.



"We have a large number of machines to manufacture our main product and now we want to find out whether there are ***common groups*** *of* ***configuration profiles*** *based on the many configuration parameters which are mostly set subjectively by our machine operators. (…)"*

"Cluster analysis is the task of ***grouping objects*** *together in such a way that objects in the* ***same group*** *are* ***similar*** *to each other and objects in* ***different groups*** *are* ***dissimilar*** *to each other. (…)"*

"Classification is the task of ***identifying*** *to which of a* ***set of categories*** *a new observation* ***belongs****, on the basis of a training set of data containing observations whose* ***category membership*** *is* ***known****. (…)"*

Entity	X1	X2	…	Xn
Δ1	Low	Low	…	42
Δ2	High	High	…	42
Δ3	High	Low	…	42
…	…	…	…	…
Δn	High	High	…	42

Figure 4: Intended functionality of an IAS for DM method selection (Zschech et al. 2020)

The intended functionality above allowed the derivation of several requirements, such as that a novel IAS should be able to process text data and that it disregards irrelevant noise. Such specific requirements were subsequently related to more generic, theory-driven meta-requirements. In particular, the design requirements of generic decision support systems (DSS) were taken from Meth et al. (2015) as prior knowledge to inform the design of the intended artifact. Table 4 summarizes the results, where R1, R2, and R3 denote meta-requirements.

ID	*(Meta-) Requirement*
R1	Increase decision quality by providing advice with high advice quality
R1.1	The system shall select DM methods with higher accuracy than guessing
R1.2	The system shall be able to remove noise from user inputs
R2	Reduce the human decision maker's cognitive effort by providing decision support
R2.1	The system shall provide the user with the ability to enter natural-language and domain-specific text
R2.2	The system shall be able to extract context and central constructs from user inputs
R3	Minimize system restrictiveness by allowing users to control the strategy selection
R3.1	The system should provide the user with the ability to review transparent assessment scores for DM method selection
R3.2	The system shall be able to operate on small amounts of text

Table 4: Summary of (meta-) requirements (Zschech et al. 2020)

3.2 Design Principles and Design Features

Based on the derived requirements, suitable methods and technologies were sought that could be incorporated into an adequate system design. More specifically, a literature review was carried out (Boell and Cecez-Kecmanovic 2014), in which various methodical approaches from the field of text mining and natural language processing were identified. The results included, for example, text classifiers, embeddings, topic models, keyword extractors, and different kinds of preprocessing techniques (Aggarwal and Zhai 2012).

These methodical approaches had to be combined and transferred into multiple processing pipelines for testing and evaluating alternative system architectures. In the sense of the DSR methodology, the concrete implementations can be understood as *design features*, upon which the system design is instantiated. The generalization of the design is then encapsulated by *design principles*, which allow an abstraction from the technical details of the solution and thus provide prescriptive knowledge for the design of a class of systems (Meth et al. 2015; Morana et al. 2019).

In order to support the mapping task[3] in an automated manner, two central aspects had to be considered. On the one hand, it had to be ensured that the IAS is capable of automatically processing natural language requests in their entirety to assign them to a class of DM methods. This step could be technically realized with the help of general text classification methods (Kowsari et al. 2019). On the other hand, it had to be ensured that the IAS automatically extracts context from the problem descriptions in the form of central constructs (e.g., keywords, phrases) that signal a match or at least a similarity between domain-specific problem descriptions and generic DM method descriptions. This step could be technically realized by using different embedding models from the field of deep learning (e.g., Bojanowski et al. 2017; Iyyer et al. 2015).

Another central aspect was to construct a suitable learning base upon which the methods above could operate to enable the system's inference. In order to treat the mapping problem as a classical supervised learning task, a large amount of training data, ideally in the form of pre-classified problem descriptions, is required. However, labeled problem descriptions from practice are only sparsely available since companies usually do not store such information in central repositories. Therefore, an alternative approach had to be developed by crawling and augmenting texts from academic articles that describe the application of DM methods (Vainshtein et al. 2018). In this way, a sufficiently large corpus could be created in an economically feasible manner.

In summary, the design of the IAS was expressed by three design principles that are concretized by four design features. In their composition, they contribute to the coverage of all previously identified design requirements. Figure 5 summarizes the relationships between the (DSS) design requirements, the design principles, and the design features.

---

[3] Please note that, in contrast to the eventually developed artifact, the original draft included an additional functionality besides the realization of the mapping task. Thus, the IAS was supposed to extract semantically relevant domain entities from the problem descriptions and translate them into the corresponding output views (Zschech, Heinrich, Horn, et al. 2019). However, this functionality was disregarded in subsequent design cycles to keep the complexity of the study manageable.

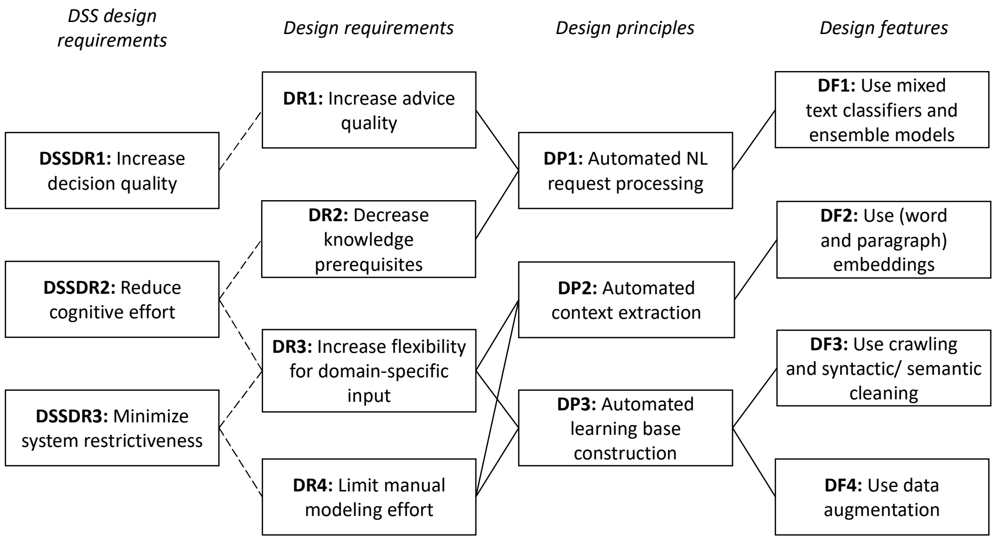


Figure 5: Relation between design requirements, principles, and features (Zschech et al. 2020)

3.3 Prototypical Instantiation and Evaluation

For the evaluation of the system design artifact, an assessment had to be performed as to how well the developed IAS is able to establish a correct mapping between real-world problem descriptions and a particular selection of DM method classes. For this purpose, the scope was limited to a selection of three predominantly employed classes: *clustering*, *prediction*, and *frequent pattern mining* (Tsai et al. 2014). This limitation was also applied when constructing the learning base in the development step. For each of the three method classes, 20 suitable problem descriptions were collected, either from industrial DSA projects or from several DM competition websites.

This collection was used for two evaluation steps. First, it was applied to determine which of the alternative processing pipelines performed best in order to transfer them as a concrete design instantiation into a prototypical implementation. Afterwards, the 60 problem descriptions were used for an external evaluation to assess the usefulness of the artifact against various reference items. These items included (i) *random guessing* as the lowest performance limit, (ii) a *novice assessment* with 20 DSA students at the beginning of their DM education, and (iii) a *baseline configuration* of the IAS in contrast to a *full configuration*. Concerning the last item, the idea was to incrementally activate individual design principles in order to measure their effects separately (Meth et al. 2015). Thus, the baseline configuration consisted only of the learning base and a set of standard text classifiers (*). Table 5 summarizes all items as part of the evaluation study.

Evaluation item	*Description*	*Design principles*	*Role within hypotheses*
Random guessing	Discrete uniform distribution	No DP	Reference item for H1
Novice assessment	DSA student survey	No DP	Reference item for H2
IAS baseline configuration	Learning base + standard text classifiers	DP3 + DP1(*)	Reference item for H3
IAS full configuration	Learning base + embeddings + advanced text classifiers	DP3 + DP1 + DP2	Test item for H1, H2, H3

Table 5: Reference and test items of the evaluation study (Zschech et al. 2020)

Corresponding to the different reference items, three design hypotheses were proposed to assess the design artifact's usefulness. In summary, the hypotheses covered the assumptions that a method selection based on a full IAS design configuration achieves higher advice quality than (i) a selection based on random guessing (H1), (ii) a selection based on the judgment capacity of DM novices (H2), and (iii) a selection based on a baseline configuration (H3).

For performance comparison, the 60 problem descriptions were classified by each evaluation item while calculating different quality metrics. When measuring overall accuracy as the proportion of correctly classified cases among the total number of cases, it was revealed that the full IAS design configuration based on all three design principles dominates all three reference items. In detail, 54 problem descriptions were assigned correctly to one of the three DM method classes, reaching an accuracy of 90%. In comparison, the baseline configuration only achieved 58%, which was still slightly higher than the mean accuracy obtained by the novices' judgment (55%). Given the setting of three DM method classes, random guessing was set to a score of 33%, constituting the lowest limit of desired advice quality.

Moreover, for hypothesis testing and to provide more stable statements about inter-group differences, confidence scores were calculated for each decision. These scores express how "sure" an algorithm is about a decision. In this way, a two-stage analysis could be conducted, including a robust version of ANOVA (Wilcox 1989) and a post hoc independent *t*-test with Bonferroni adjustment. While the ANOVA returned a significant result for the overall test that at least two evaluation items were different, the *t*-tests returned significant results on H1 and H2 at the 0.01 level, and on H3 at the 0.05 level. These results support the three hypotheses and confirm that an IAS based on all design principles indeed increases the advice quality using natural language problem descriptions. Table 6 summarizes the results of the *t*-tests.

Hypothesis	*Level*	*versus Level*	*Difference*	*p-value*
H1	Full configuration	Random guessing	0.369	< .0001*
	Baseline configuration	Random guessing	0.264	< .0001*
	Novice assessment	Random guessing	0.219	< .0001*
H2	Full configuration	Novice assessment	0.147	0.0006*
H3	Full configuration	Baseline configuration	0.102	0.0165*
	Baseline configuration	Novice assessment	0.044	0.2968

Table 6: Post hoc *t*-test results for hypotheses H1–H3 (Zschech et al. 2020)

4 Taxonomic Framework for Method Evaluation

Research objectives		
RO3	Development of a taxonomic framework for method evaluation	
3a	Survey of prognostic solutions using public benchmark data	
3b	Development of a taxonomic evaluation framework	
3c	Reconstruction of methods and application of the evaluation framework	
Reference to original work		
Zschech, Bernien, et al. (2019)	Publication P4	Appendix II: D

Table 7: Research summary for Chapter 4

This chapter is concerned with the development of a taxonomic evaluation framework for the systematic assessment of data-driven methods. Inspired by the research approach conducted in Chapter 2, it was observed that taxonomies serve as a viable tool to decompose multi-layered objects or entities into their inherent parts and facets. Concerning the decomposition of data analysis problems, the distinction between *analysis objectives*, *data properties*, and *analysis methods* proved to be an adequate way to develop a comprehensive systematization framework. Beyond that scope, however, discussions with experts from research and industry revealed the potential to expand such a framework to include further dimensions. Taking the structure of classical DM procedure models, such as CRISP-DM (Mariscal et al. 2010), conceivable extensions covered dimensions related to *data pre-processing* and *evaluation*. In return, however, a smaller focus had to be set to keep the variability of such additional dimensions manageable. These steps led to the creation of a new method proposal for the development of taxonomic evaluation frameworks. While the overall composition can be considered an innovative contribution, the core components consist of methodological steps derived from well-established research approaches. The general procedure of the method proposal, as well as an instantiated example, are summarized in Figure 6.

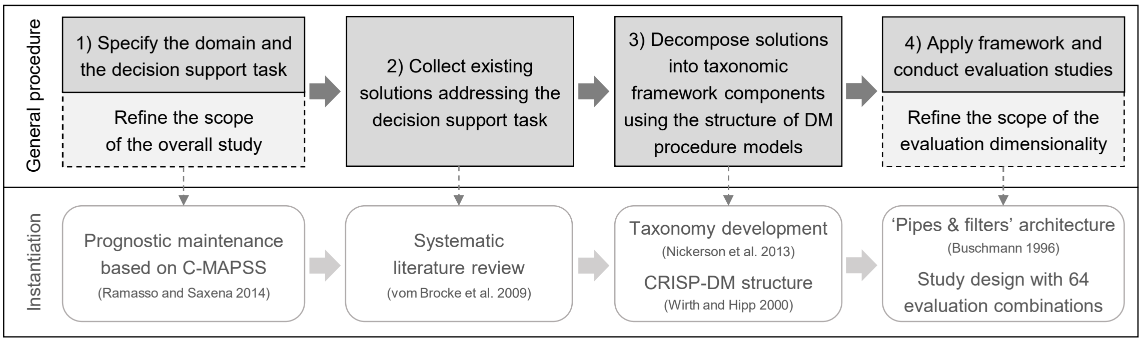


Figure 6: Method proposal for the development of taxonomic evaluation frameworks

In the **first step**, the domain and the decision support task have to be specified. The task must be sufficiently well delimited and needs to allow for support from data-driven methods that can be evaluated using quantitative evaluation metrics such as diagnostic or predictive decision tasks.

Optionally, the scope of the overall study can then be further refined to focus on particular settings or circumstances. For the exemplary instantiation, the field of maintenance is chosen as the domain, focusing on the task of prognostic decision support. Moreover, the scope is refined by considering a turbofan engine degradation setting based on NASA's C-MAPSS scenario (Ramasso and Saxena 2014) to keep the study's complexity manageable.

In the **second step**, existing solution approaches based on data-driven methods have to be collected that address the specified decision support task. In this way, an overview can be obtained about the alternative design options for building data-driven solutions. For the realization of this step, it is advisable to draw on established research methods for conducting a systematic literature review (Webster and Watson 2002). In the demonstration example, the guidelines proposed by vom Brocke et al. (2009) were adopted.

In the **third step**, the identified solutions have to be decomposed into modular components to obtain the taxonomic structure of the evaluation framework. For this step, it is advisable to adopt the guidelines proposed by Nickerson et al. (2013), as already introduced in Chapter 2. However, as mentioned above, the extraction of dimensions and characteristics is supposed to follow the general structure of DM procedure models, which are basically organized into the steps of *domain understanding*, *data understanding*, *data pre-processing*, *modeling*, and *evaluation* (Mariscal et al. 2010; Wirth and Hipp 2000).

In the **fourth step**, the evaluation framework is applied, and quantitative studies are conducted by reconstructing the identified solution components for different contexts. In this way, the extracted framework elements serve as evaluation options that are iteratively modified under ceteris paribus conditions. Thus, by using a "pipes and filters" architecture (Buschmann 1996), all conceivable combinations of pre-processing and prognostic modeling methods can be studied based on different data properties concerning their impact on multiple evaluation criteria. However, instead of using the entire evaluation framework, the option should be considered to refine the scope of the study design to focus on specific aspects. Such an option is also chosen in the demonstration example by focusing on 64 evaluation combinations.

In the next sections, the application of the proposed method is demonstrated in further detail. Thus, by focusing on *prognostic maintenance solutions,* the remaining structure follows the composition of the sub-objectives 3a–3c (cf. Table 7). Please note that the results of 3a and 3b are already covered in full detail by publication P4. By contrast, the quantitative evaluation results of 3c have not yet been part of any publication.

4.1 Survey of Prognostic Solutions

Due to the scarce availability of run-to-failure data in industrial environments, the development of prognostic maintenance solutions is primarily based on synthetic data collections. For this purpose, there have been several initiatives to generate public benchmark datasets based on laboratory experiments and advanced simulations. Such initiatives cover a variety of technical settings including milling machines, bearings, turbofan engines, and battery charging cycles (Eker et al. 2012; Lei et al. 2018). Among these examples, the turbofan scenario based on NASA's

C-MAPSS data is one of the most dominating benchmark scenarios in the prognostics community. Due to their realistic properties in terms of high-dimensional sensor measurements and masked fault effects (cf. Chapter 2.3), the C-MAPSS data have already been used by hundreds of researchers from various disciplines, bringing forth a wide variety of prognostic solution approaches (Ramasso and Saxena 2014). For this reason, the study's scope was explicitly limited to this specific scenario as it provides an extensive knowledge base while being sufficiently manageable when assessing individual solution approaches in more detail.

To identify the large amount of studies developing C-MAPSS-based prognostic solutions, the review guidelines proposed by vom Brocke et al. (2009) were followed. More specifically, this included (i) a conceptualization of the topic to retrieve appropriate search terms, (ii) a database search using several digital libraries, (iii) a forward and a backward search based on relevant key contributions, and (iv) a specification and application of filter criteria to remove irrelevant literature from further analysis. In this way, it was possible to obtain 227 unique hits before applying filter criteria (day of search: 24-09-2018). After filtering, the number of items was reduced to 106 relevant studies[4].

4.2 Taxonomic Evaluation Framework

In the next step, the vast corpus of C-MAPSS studies was used to develop the structure of the evaluation framework. Following the guidelines proposed by Nickerson et al. (2013), the development process was structured into several steps and iterations, similar to the procedure in Chapter 2. The meta-characteristic was defined as characteristic components of data-driven prognostic solutions. Concerning the ending conditions, most suggestions from the authors could be adopted without significant changes as they provide a solid basis to determine the end of the iterative process. After specifying those properties, the actual step of extracting dimensions and characteristics was carried out. At this stage, the procedure proposed by Nickerson et al. (2013) was refined by additionally taking into account the general structure of the CRISP-DM procedure model (Wirth and Hipp 2000) to distinguish between characteristic components of data-driven solutions. As a result, it was possible to identify (i) two dimensions related to *domain* and *data understanding*, (ii) four dimensions related to *pre-processing*, (iii) one top-dimension and several intangible sub-dimensions related to *modeling*, and (iv) one dimension related to *evaluation*.

Moreover, as specified by Nickerson et al. (2013), the extraction process covered both empirical as well as conceptual knowledge. Empirical knowledge was directly obtained when analyzing each individual study in the corpus and extracting elemental parts of prognostic solutions. Conceptual knowledge, on the other hand, was derived from existing survey papers and systematizations that were identified during the literature review above (e.g., Ramasso and Saxena 2014; Saxena, Celaya, et al. 2008). In this way, it was possible to use prior expert knowledge and organize empirical observations into pre-defined categories.

The results of the taxonomy development are summarized in Table 8. The derived elements can be considered as design options when implementing data-driven prognostic solutions in similar

---

[4] Full list of references: https://www.researchgate.net/publication/335611604 (last access: 01-06-2020)

settings. While the first two dimensions specify the context in which different data-driven methods based on various pre-processing and modeling components can be tested, the evaluation dimension covers multiple options for assessing the quality of the results. For a more detailed description of the dimensions and characteristics, please refer to the full study (Zschech, Bernien, and Heinrich 2019).

CRISP-DM	Dimensions	Characteristics		
Domain & data understanding	Fault modes	Single fault mode		Multiple fault modes
	Operational conditions	Single condition		Multiple conditions
Pre-processing	Normalization	Standardization		Rescaling
	Noise reduction	Moving average	Exponential smoothing	Polynomial smoothing
	Feature selection	Manual selection	Filter	Wrapper
	Dimensionality reduction	Hierarchical		Non-hierarchical
Modeling	Prognostic approach	Direct RUL-mapping	Indirect RUL-mapping via HI	Similarity-based matching
Evaluation	Performance metric	Accuracy-based	Precision-based	Prognostic-specific metric

Table 8: Taxonomic evaluation framework based on C-MAPSS studies

4.3 Exemplary Framework Application

After the extraction of the framework structure, the derived elements can be used to create a study design for different evaluation purposes. This step is demonstrated below by taking selected characteristics for each framework dimension and implementing them with concrete approaches. Figure 7 summarizes the selected elements of the exemplary study design. Please note that in the given scenario, some dimensions can be skipped, which is possible, for example, for all four pre-processing dimensions.

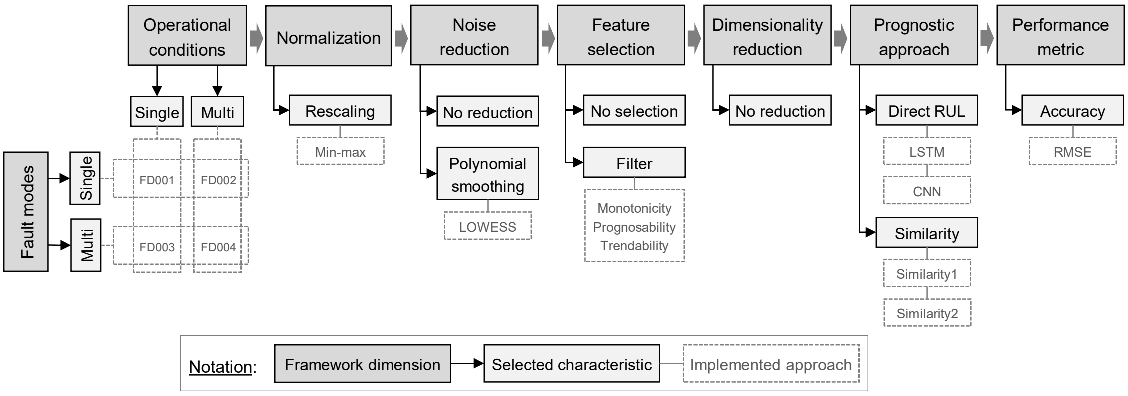


Figure 7: Exemplary study design derived from the evaluation framework

In order to consider different complexity levels of the domain, four alternative datasets of the C-MAPSS collection are chosen. In particular, FD001 and FD002 are used as they represent scenarios with a single fault mode while differing in the number of operational conditions. Likewise, FD003 and FD004 are used to consider multiple faults combined with different operational conditions.

Concerning the construction of the pre-processing pipeline, a *normalization* step is realized by using a rescaling approach through a min-max transformation (Tao et al. 2016). Subsequently, in a first variant, *noise reduction* is implemented via locally weighted scatterplot smoothing (LOWESS) as a concrete approach for polynomial smoothing (Khelif et al. 2017). In a second variant, the step of noise reduction is skipped to examine the particular impact on the overall performance. A similar approach is carried out for the step of *feature selection*. In a first path, all input features are used without any selection procedure. In a second path, a filter method is applied based on a weighted combination of the metrics “monotonicity”, “prognosability”, and “trendability” (Coble 2010). The next step of *dimensionality reduction* is skipped without any particular implementation.

For the *prognostic modeling* step, the two categories of direct RUL-mapping and similarity-based matching are chosen. The direct RUL-mapping is realized with two different kinds of deep neural networks. More specifically, a long short-term memory (LSTM) network (Zheng et al. 2017) and a convolutional neural network (CNN) (Babu et al. 2016) are implemented. The similarity-based approach is also realized through two specific implementations. While both of them share the same procedure for constructing health index (HI) curves (Khelif et al. 2017), they differ in the applied approach for curve fitting and the type of similarity score (Malhotra et al. 2016; Wang et al. 2017). Finally, for *performance evaluation*, the root mean square error (RMSE) is used as a standard accuracy-based metric to assess the quality of the RUL estimation task (Lim et al. 2016).

The implementation[5] of the individual approaches described above is organized in modules using the programming language Python. The general structure of the taxonomic evaluation framework allows modules from different framework dimensions to be stacked in sequential processing steps using a “pipes and filters” architecture (Buschmann 1996). In this way, modular pipelines can be constructed in which the output of one module represents the input of the subsequent one. For this purpose, a dictionary is created to check the combinability of different modules with each other. In the present example of the C-MAPSS scenario, the developed framework allows the combination of all dimensions without any restrictions, so that a fully populated evaluation matrix can be obtained. However, it is also conceivable that some cells of the matrix remain unoccupied in the case of limited combinability. To automatically generate the evaluation results, conditional statements are used to execute those modules that correspond to a particular combination, while all predefined combinations are executed using loop constructs.

For demonstration purposes, the resulting evaluation matrix is illustrated in Table 9. The framework dimensions and the implemented approaches cover row and column elements, while the cells of the matrix reflect the results of the chosen evaluation metric. For better readability, the evaluation matrix is organized into four quadrants according to the datasets FD001–FD004 covering the different complexity levels of the scenario. Pre-processing alternatives are reflected by columns, while alternative prognostic models are organized in rows. An additional color scheme, adjusted for each quadrant, highlights the differences in performance. The lower the RMSE values, the stronger the color intensity, indicating that an individual evaluation pipeline performs better than another.

---

[5] Further details on each implemented approach, such as the choice of hyperparameters, can be found in Appendix I: Implementation Details.

Metric: RMSE		**Single operational condition**				**Multiple operational conditions**			
		No noise reduction		Polynom. smoothing		No noise reduction		Polynom. smoothing	
		No select.	Filter	No select.	Filter	No select.	Filter	No select.	Filter
Single fault mode	LSTM	15.19	16.02	13.85	15.40	32.15	32.40	30.68	31.68
	CNN	15.23	17.77	14.86	17.31	30.67	30.64	30.67	30.58
	Similarity1	18.37	19.21	19.85	19.95	29.55	29.84	28.77	28.56
	Similarity2	14.42	16.31	16.03	18.37	23.88	24.44	24.32	24.38
Multiple fault modes	LSTM	18.84	18.59	17.81	32.47	34.56	38.41	33.59	39.52
	CNN	18.20	22.83	15.83	25.32	31.79	32.49	32.36	32.72
	Similarity1	28.57	27.89	29.90	30.22	32.55	33.41	33.42	33.93
	Similarity2	20.31	22.25	22.52	22.74	27.24	27.36	27.18	27.94

Table 9: Evaluation results for selected framework elements

By using the resulting evaluation matrix, it is possible to draw several conclusions about the suitability of alternative data-driven methods in different settings. For example, it can be observed that direct prognostic models based on deep neural networks (i.e., LSTM and CNN) tend to perform slightly better than similarity-based approaches in settings with single operational conditions, especially when multiple fault modes are present. By contrast, similarity-based models tend to perform better than direct approaches in scenarios with multiple operational conditions. This observation is particularly true for the second similarity approach (Similarity2), which, however, generally shows high accuracies across all settings.

Simultaneously, it is possible to assess the adequacy of combining particular method components. For example, it can be noted that neural networks without explicit feature selection, in most cases, achieve much better results compared to their variants with feature selection using the filter approach. This observation confirms the assumption that deep neural networks are generally capable of automatically extracting relevant features without the need for additional feature engineering (LeCun et al. 2015). Similarly, it can be noted that polynomial smoothing, except in the case of FD002 (i.e., single fault, multiple operational conditions), generally reduces the performance of similarity-based approaches. One explanation could be that noise reduction removes essential information from the signals that would have been relevant for matching similar curve segments. Therefore, such method combinations should be avoided in comparable settings.

Overall, the few analysis examples illustrate which useful insights can be gained by applying such a taxonomic evaluation framework. For demonstration purposes, the scope has been kept deliberately small, so even more dimensions, characteristics, and concrete implementations are conceivable to expand the scope and conduct more in-depth analyses. Furthermore, neither the developed framework derived from the C-MAPSS studies nor the overall method proposal for constructing the framework is restricted to the specific case at hand. Instead, it is feasible to apply both approaches to other settings, which will be further discussed in Chapter 6.

5 Method Application Under Industrial Conditions

Research objectives	
RO4	Application of a prognostic method under industrial conditions
4a	Conceptualization of a novel solution to address missing label situation
4b	Prototypical implementation and evaluation of the solution approach
Reference to original work	
Zschech, Heinrich, Bink, et al. (2019)	Publication P5 — Appendix II: E

Table 10: Research summary for Chapter 5

In contrast to laboratory settings and simulations, as in the case of the C-MAPSS scenario, it is a considerable challenge in real production environments to detect and anticipate critical machine behavior in a proactive manner. Often there is a lack of knowledge about thresholds and tolerance limits that mark necessary points of intervention. Moreover, in many cases, machines are operated and maintained with great caution, so that actions are taken long before necessary interventions are required. From a prognostic point of view, this situation is often referred to as a "missing label" problem, which can be seen as a significant hurdle in the development of predictive decision models (Gouriveau et al. 2013).

Against this background, the present chapter deals with the fourth research objective and addresses the application of a prognostic method under industrial conditions. For this purpose, a maintenance scenario of a German car manufacturer is considered as an exemplary case. More specifically, the scenario refers to a milling machine with replaceable milling tools that are subject to natural wear and tear. In order to reduce the wear effect, imperfect corrections have to be carried out by machine operators until the milling tools finally have to be replaced. Although extensive sensor data are captured during the production process, there are no thresholds specified indicating when a tool replacement should ideally be carried out. Instead, the operators' decisions regarding tool replacements are exclusively based on (i) their perception during visual tool inspections, (ii) their empirical knowledge, and (iii) their individual risk preferences. Thus, less experienced machine operators with more risk-averse attitudes tend to replace tools well before the actual end of useful life. In contrast, risk-taking machine operators tend to carry out late replacements, risking impaired product quality. Overall, this leads to inefficient use of resources, which is why a proactive solution approach for better decision support is required.

In order to carry out this kind of research and develop a novel solution approach, a data science study was conducted by following the general steps of DM procedure models (Mariscal et al. 2010). More specifically, concerning the sub-objectives in Table 10, the solution approach was first conceptualized on an abstract level (4a). Subsequently, the solution was prototypically implemented and evaluated using real data collections provided by the case study partner (4b). In the following sections, the results of both sub-objectives are briefly described.

5.1 Conceptualization of a Solution Approach

Due to a lack of objective information on when a tool replacement should be carried out, machine operators' decisions are made on a subjective basis taking into account individual risk preferences. To address this problem, a prognostic decision model had to be created that provides a reference point by predicting the milling tools' RUL in order to reduce subjectivity in the decision process. However, the lack of objective information also implied the absence of adequate labels, which were required for learning a suitable prognostic model. In other words, if a prognostic model had been trained based on all previous observations, the model would have only reflected the decisions of the machine operators and not the technically possible RUL of the milling tools.

In response, the core idea of a novel solution approach was to separate "good decisions" from "bad decisions" based on latently available information hidden in historical data records about executed tool replacements. For this purpose, the problem space was conceptualized using two orthogonally related dimensions. The first dimension refers to the *time* when a tool replacement was carried out, distinguishing between early and late replacements. The second dimension refers to the *condition* of a milling tool, distinguishing between damaged and undamaged tools. Even if this information was not directly available in the data, it was reasonable to assume that a critical damage pattern must also be reflected in the recorded sensor values of the milling machine. By separating the two levels in both dimensions, a four-field matrix can be set up as illustrated in Table 11. On this basis, it is possible to differentiate between four types of tool replacements due to subjective decisions:

- **Type 1** represents undamaged tools that have been replaced correctly at a late time, implying an efficient use of resources.
- **Type 2** represents damaged tools that have not been replaced in time, leading to impaired product quality.
- **Type 3** represents undamaged tools that have been replaced too early, resulting in high tool costs and truncated data for model training.
- **Type 4** represents damaged tools that have been replaced correctly at an early time, also corresponding to efficient use of resources.

		Condition	
		Tool undamaged	Tool damaged
Time	Replacement late	Type 1 – GOOD (efficient tool usage, type 3 prevented)	Type 2 – AVOID (impaired product quality)
	Replacement early	Type 3 – AVOID (high tool costs)	Type 4 – GOOD (efficient tool usage, type 2 prevented)

Table 11: Four-field matrix for the distinction of tool replacements (Zschech, Heinrich, Bink, et al. 2019)

In order to ensure resource-efficient replacements in productive use as illustrated by types 1 and 4, a maintenance system had to be established consisting of two analytical components. The first component is a diagnostic decision model that continuously checks whether a milling tool shows any signs of imminent damage. If this is the case, it has to be replaced. If this is not the case, a prognostic decision model trained on type 1 observations is used to determine the RUL of the tool, since type 1 observations represent tools that have been correctly replaced at a late stage. This procedure is associated with the assumption that those replacements are close to the actual end of useful life based on the empirical knowledge of more experienced machine operators.

5.2 Prototypical Implementation and Evaluation

For the implementation of the solution approach, a systematic data science study was conducted, following the steps of domain and data understanding, data preparation, modeling, evaluation, and deployment (Mariscal et al. 2010). More specifically, the scope of the implementation was primarily limited to the distinction of tool replacements into the four types described above and the development of a prognostic model. The development of a diagnostic model, on the other hand, was only partially addressed as it required more profound system knowledge, which was not attainable at the time of the implementation. Figure 8 summarizes the implemented solution approach and highlights relevant case characteristics and applied methods.

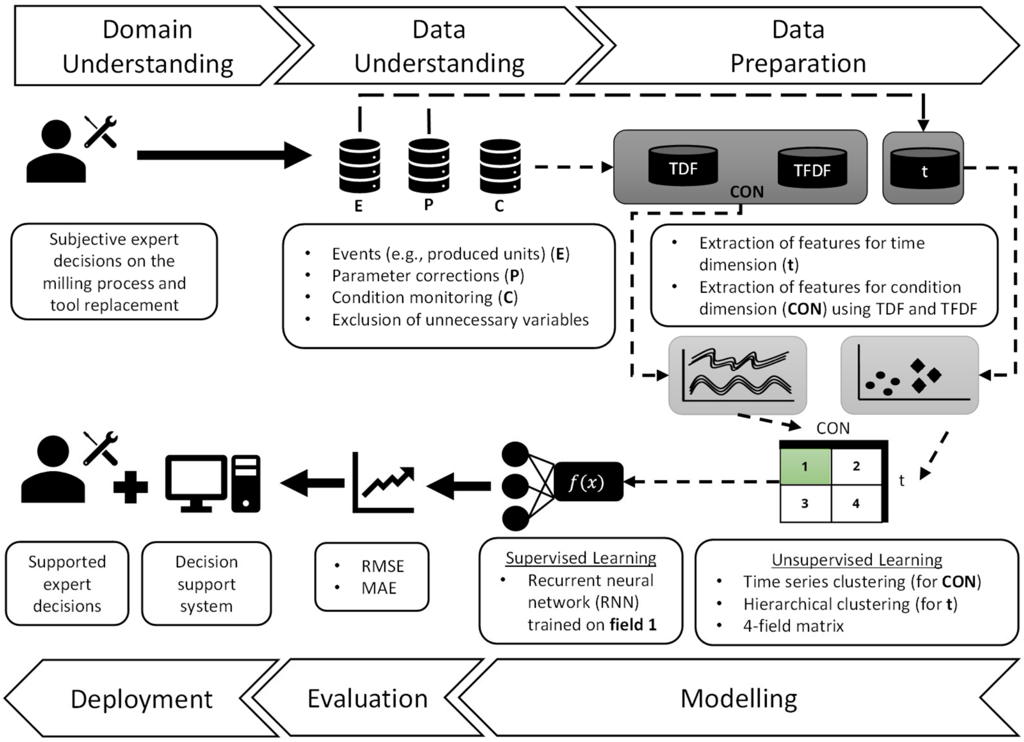


Figure 8: Implemented solution approach (Zschech, Heinrich, Bink, et al. 2019)

After establishing a domain understanding, as already illustrated during the conceptualization of the solution approach, the step of data understanding was carried out. For this purpose, the case study partner provided a representative dataset containing information on an output volume of 88,125 processed parts. During the processing of these parts, a total of 67 tool replacements were recorded. More specifically, the dataset contained information about (i) recorded events such as processed units or cycle times, (ii) applied parameter corrections representing imperfect maintenance actions, and (iii) condition monitoring data reflecting measurable machine behavior at a certain point in time (cf. Chapter 2.3).

In the next step, the data collections were processed in order to use them for subsequent modeling tasks. Thus, following the structure of the four-field matrix described above, characteristic features for both dimensions had to be selected and prepared accordingly. In particular, event records and parameter corrections served to derive features for the *time dimension*. The time-series signals from condition monitoring, on the other hand, were used for the *condition dimension* by extracting time-domain features (TDF) and time-frequency domain features (TFDF) (Goyal and Pabla 2015).

After that, the actual modeling step was carried out. This step included the two successive tasks of (i) detecting structural patterns in all recorded observations to assign them to the four-field matrix, and (ii) developing a prognostic model based on representative observations. For the first task, methods from the field of unsupervised ML were applied (Everitt et al. 2011). More specifically, an agglomerative hierarchical clustering approach (Sneath and Sokal 1973) was implemented to separate observations of the *time dimension* into early and late replacements, while a time series clustering approach was used to distinguish between observations of damaged and undamaged tools. By using the derived clusters of both dimensions, it was possible to relate them orthogonally to each other and assign the resulting four subsets to the respective quadrants of the four-field matrix. Subsequently, the prognostic task was treated as a supervised learning problem for RUL estimation (cf. Chapter 4). For this purpose, two variants of recurrent neural networks (RNN) (Williams 1995) with alternative feature sets were implemented using type 1 observations as training instances to develop a prognostic model. Moreover, three different prediction horizons were chosen in order to estimate the RUL of the milling tools in the short, medium, and long term.

For the evaluation of the prognostic models, two performance metrics were applied: RMSE and mean absolute error (MAE) (Pan et al. 2014). The results showed that the RNNs were able to adequately learn the regularities of the time series, as they achieved small estimation errors for all three forecasting horizons. For example, having an average lifetime of 1,315.3 processed units per milling tool, the best performing model under- or over-estimated tool lifetime by an average of 82, 80, and 77 units for the prediction horizons t+35, t+175, and t+350, respectively.

Finally, in a simulated deployment step, it was further examined which advantage the prognostic model would provide if it were applied in operational processes. For this purpose, the model was used to estimate the RUL for type 3 observations in which tool replacements were performed too early. By comparing the actual tool lifetime with the models' RUL estimates, it was possible to quantify the unused service life. As a result, it could be observed that it would have been possible to save about 4–5 milling tools within the period under consideration. Having a total number of 67 tools, this corresponds to cost savings of approximately 6–7%.

6 Discussion of the Results

This chapter offers a discussion of the achieved results. To this end, all individual artifacts have been critically reflected in the respective publications, P1–P5, concerning (i) merits and limitations as well as (ii) implications for further research and practical applications. In the following discussion, several of these aspects are taken up again and considered at a more cohesive level. More specifically, this involves a consideration of connections between the developed artifacts and related work, as well as the generalization and transferability of the achieved results. The findings of the chapter can be regarded as a research agenda and outlook for subsequent work.

6.1 Connections Between Developed Artifacts and Related Work

From a joint consideration of all four focus areas of this thesis, there are several connections between the individual artefacts as well as relations to related work. Figure 9 provides a summary of relevant connections, which are briefly discussed below. Continuous arrows represent connections that have been explicitly considered in this work, while dashed arrows indicate research opportunities for future projects.

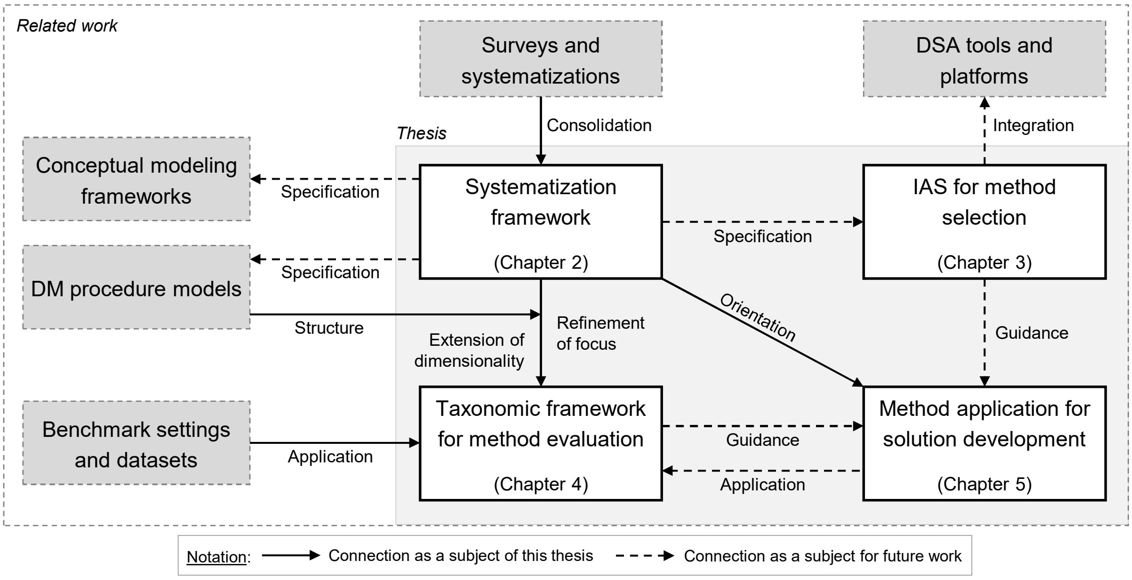


Figure 9: Connections between developed artifacts and related work

The systematization framework from Chapter 2 is the result of consolidating different classification schemes derived from academic surveys, as well as expert knowledge from industry, into a structured framework. The framework provides a viable instrument to decompose complex data analysis problems into single dimensions and corresponding characteristics. Therefore, it can serve as a tool for communication purposes to create a shared understanding between different stakeholders involved in multidisciplinary DSA projects, such as domain experts, analysts, and IT professionals. In this context, the exemplary application of the framework in Chapter 2.3 has demonstrated how data analysis scenarios with their central properties can be described systematically to provide an orientation for method application and solution development in similar settings as that one described in Chapter 5. As a result, the solution space can be limited to the essential properties, and a reusable template can be defined instead of rediscovering a problem solution every time from scratch.

At this point, some open connections to related work can be identified (cf. Chapter 1.3). Initially, the author of this thesis pursued the goal of using the systematization framework to create a reusable knowledge base in the sense of reference models and composite solution models, following the examples of Eckert and Ehmke (2017) and Brodsky et al. (2015). Therefore, a new modeling approach should be developed that allows the creation of predefined solution templates analogous to design patterns. In other words, the dimensions and characteristics of the framework should be related to each other within concrete maintenance solutions that describe standard templates to address recurring DSA tasks. Nevertheless, due to further feedback cycles with practitioners, this modelling approach was discarded during the dissertation project. Thus, it was not possible, for example, to agree on a suitable level of abstraction that ensures a sufficient degree of problem specificity while allowing a high degree of generalizability and reusability.

However, only recently the work by Nalchigar et al. was discovered in this context. The authors pursue a very similar idea and propose a comprehensive conceptual modeling framework for the development of DSA solution patterns (e.g., Nalchigar et al. 2019; Nalchigar and Yu 2020). Thus, the authors' extensive modeling efforts could be applied in combination with the derived elements of the systematization framework to continue to idea of creating standardized solution templates for the domain of data-driven maintenance. Although no concrete cooperation with the authors has yet been established, this attempt will be taken up again in future projects.

A similar connection is also conceivable to DM procedure models, which could be enriched with more domain specificity by integrating the dimensions and characteristics of the systematization framework. For this purpose, initial exchanges with the authors of the DMME model (data mining methodology for engineering applications) (Huber et al. 2019) have already taken place to enhance their industry-tailored procedure model with maintenance-related specifications.

Beyond the potential to develop DSA solution patterns and refine DM procedure models, the systematization framework served as a basis for a new method proposal towards the development of taxonomic evaluation frameworks, as demonstrated in Chapter 4. This approach required placing a smaller focus on specific decision support tasks, while the dimensionality of the framework was extended by considering the general structure of DM procedure models. Throughout an exemplary instantiation of the method proposal in Chapter 4.3, it has been demonstrated how prognostic maintenance solutions can be decomposed into modular parts to retrieve a variety of alternative design options. On this basis, more fine-grained evaluation studies were possible in order (i) to assess the suitability of alternative design options for different contexts, and (ii) to verify the adequacy of combining particular solution components.

For demonstration purposes, the exemplary instantiation was based on NASA's turbofan engine degradation scenario, which is known as a common benchmark setting in the prognostics community. However, it is also possible to apply such a taxonomic evaluation methodology to completely new scenarios from industrial practice that have not emerged as typical benchmark settings based on ideal run-to-failure records. Instead, other challenges could be of particular interest, such as those posed by the central properties of the milling scenario from Chapter 5 (e.g., missing labels, imperfect maintenance). Conversely, the results of the evaluation methodology can then offer guidance for the application of suitable data-driven methods under similar circumstances in order to support the systematic development of prognostic solutions instead of going through many trial-and-error cycles.

While the type of guidance provided by the evaluation framework from Chapter 4 is primarily directed at target groups with more advanced DSA experience (e.g., programmers, researchers), DSA novices require guidance at a more abstract level. For this purpose, the developed artifact described in Chapter 3 provides a novel IAS that takes problem descriptions expressed in natural domain-specific language as input and offers advice regarding the most suitable class of methods to address the problem. In this way, the artifact can assist novices at the beginning of their DSA projects as an entry point to obtain a better understanding of possible solution directions as well as necessary foundations for the application of the recommended class of methods.

Concerning possible connections to related work, the proposed IAS can either be used as a standalone application or as a novel add-on embedded into existing DSA platforms such as RapidMiner or KNIME (cf. Chapter 1.3). To this end, it is also conceivable to combine it with other types of assistance systems in order to provide further guidance to novice users as soon as an adequate class of methods is determined.

Moreover, in the current version, the proposed IAS works on a general level without any particular domain focus. As such, it remains an open research topic to use the findings of the systematization framework from Chapter 2 to focus on recurring data analysis problems in industrial maintenance and evaluate the usefulness of the IAS in that specific domain. This point is further discussed in the following section when considering the transferability of the achieved results.

6.2 Generalization and Transferability of the Results

The developed artifacts differ to some extent in their scope, and therefore the resulting contributions and achieved results can be positioned at different *levels of applicability*. Correspondingly, it is possible to identify several limitations of the thesis as well as associated prospects for future research.

For this purpose, the following discussion distinguishes between four levels of applicability. The lowest level refers to a *specific case* with a corresponding problem space for which an artifact was applied or explicitly developed. The second level covers a more general *class of problems* by focusing on several properties of interest while abstracting from too specific conditions at the case level. Following this line, the third level represents the *application domain* with its characteristic peculiarities, whereby the current work mainly focuses on the domain of industrial maintenance. Finally, the highest level refers to the *general applicability* of the achieved results independent of any domain properties, problem classes, or particular case characteristics.

Table 12 summarizes the different levels of applicability for the four focus areas of this thesis. On this basis, it is possible to illustrate and discuss the generalization and transferability of the achieved results. The green cells symbolize the current focus of this thesis, while the white cells highlight prospects for future work. Additionally, the grey arrows indicate which primary research direction should be prioritized in subsequent work, i.e., whether the artifacts in question require more generalization or more specialization to achieve a higher level of maturity.

Level of applicability	**Systematization** (Chapter 2)	**Method selection** (Chapter 3)	**Method evaluation** (Chapter 4)	**Method application** (Chapter 5)
General applicability	• DSA taxonomy with tripartite meta-characteristic	• Design principals/ features of a text-based IAS	• Method proposal for framework development	-
	• Consideration of additional dimensions	• Additional design cycles (e.g., explainable AI)	• Stronger formalization	-
Application domain	• Industrial maintenance	• Unspecified	• Industrial maintenance	• Industrial maintenance
	• Manufacturing in general, social media, health care	• Manufacturing/ maintenance, social media, health care	• Computer vision, business process management	-
Problem class	• Problem classes derived from literature + interviews	• Focus on three DM method classes	• Prognostics ('run-to-failure') • Turbofan engine	• Prognostics ('missing labels') • Machine tools (milling)
	• Continuous verification with additional sources	• Extension to further method classes	• Bearings, milling machines, battery charging cycles	• Cutting, drilling, grinding, polishing
Specific case	• Application to C-MAPSS and milling case	• Not yet applied	• C-MAPSS data	• German car manufacturer
	• Application to further cases in industry	• Field study	• Additional evaluation studies • Other benchmark data	• Operational use • Other industry partners

Notation: 🟩 Current focus of the thesis ☐ Prospects for future work ⇨ Research direction for future work

Table 12: Levels of applicability of the achieved results and prospects for future work

Systematization Framework

The systematization framework from Chapter 2 generally covers a large number of problem classes, as derived from academic literature and expert interviews, and consolidates them within a structured taxonomy. In this way, the framework has been implicitly applied to an extensive number of cases. In contrast, the explicit demonstration of the framework's applicability on the case level only covered the two cases of the milling machine (cf. Chapter 5) and the C-MAPSS setting (cf. Chapter 4). Therefore, further examples of industrial cases as well as additional sources (e.g., feedback from more experts) are required to verify the robustness of the framework and examine how well the current dimensions and characteristics cover central properties of additional maintenance scenarios.

As to the broader applicability of the framework results, it can be stated that the applied procedure is not strictly limited to the domain of industrial maintenance. By following the guidelines of Nickerson et al. (2013), a sufficiently generic approach was pursued, which was only modified by choosing a tripartite meta-characteristic that reflected the general structure of recurring data analysis problems. Accordingly, this procedure can also be applied to broader contexts, such as manufacturing in general (e.g., Brodsky et al. 2015), or in completely different domains, such as social media (e.g., Kleindienst et al. 2015) and health care (e.g., Hogl 2003). Nevertheless, when discussing the results with other DSA researchers, some criticism arose that the tripartite meta-characteristic is not yet sufficient to capture the core of some data analysis problems in more complex settings. As such, it was argued, for example, that it currently lacks a more fine-grained consideration of further pre-processing dimensions. Thus, a primary research direction should be to strive for a further generalization of the approach by taking into account other potential dimensions or meta-characteristics. A conceivable extension for this purpose has already been considered when creating the taxonomic evaluation framework for the C-MAPSS scenario in Chapter 4.

Intelligent Assistance System for Method Selection

The artifact developed in Chapter 3 shows a high level of universal applicability. By abstracting from the technical details of a concrete implementation and formulating generic design principles, it was possible to derive prescriptive knowledge for the design of a class of systems that assist DSA novices in method selection. In this respect, no concrete restrictions have been made, neither at the domain level nor at the specific case level. The only restrictions can be found at the problem class level for demonstration purposes, where the selection of suitable method classes was limited to three predominantly applied DM method classes. However, due to the generic system design, an extension to further method classes is conceivable without any significant changes.

Nevertheless, the lack of specialization also entails some limitations. As such, a robustness analysis revealed that the current learning base of the proposed IAS, with mixed and unbalanced entries from multiple domains, leads to several distortions in the recommendation step. Therefore, it is planned in subsequent work to obtain a learning base that focuses on one particular domain in order to keep the domain-specific vocabulary more manageable. For this purpose, the systematization framework from Chapter 2 provides a valuable tool to specify the relevant surrounding of recurring data analysis problems in the domain of industrial maintenance. However, it is also conceivable to consider other domains such as those mentioned above. At the same time, it is intended to evaluate the artifact's usefulness by conducting field studies in cooperation with partners of a specific domain and associated validation datasets. In this way, a more realistic evaluation can be carried out since the current assessment is based on limited validation data and the judgment capacity of DSA students under laboratory conditions.

In addition to the prioritized research direction of further specialization, there is also the alternative approach of integrating additional system functionalities into the proposed IAS to guarantee a higher degree of domain independence. Particularly, an explainable AI component shall be introduced in the next design cycle to better trace and comprehend which keywords are responsible for determining a particular method class (Mathews 2019). In this way, it is expected that an increasingly robust learning base will be constructed by incrementally reducing domain-related biases when working with learning instances from multiple domains.

Taxonomic Framework for Method Evaluation

In contrast to the proposed IAS, the results of the taxonomic evaluation framework in Chapter 4 mainly concentrate on the domain of industrial maintenance. As derived from the systematization framework, the problem class was explicitly set to a prognostic decision support task and, more specifically, to run-to-failure scenarios, which are frequently applied for prognostic solution development. Furthermore, NASA's widely used C-MAPSS case was chosen on the case level, which also represents a broader class of technical settings. Thus, by applying the evaluation framework, useful insights were derived that provide not only a valuable contribution to the specific C-MAPSS case but also to the more general class of degrading turbofan engines. These insights can be further extended in future work by conducting more evaluation studies with additional framework dimensions, characteristics, and implementation variants.

Nevertheless, the specifications made at the individual levels can also be replaced by other conditions as desired. For this purpose, an attempt was made to introduce a sufficiently generic method proposal that can be applied to any kind of (i) domain, (ii) (DSA) problem class, and (iii) specific case. For example, instead of focusing on the C-MAPSS collection, similar benchmark datasets can be used for the technical setting of degrading turbofan engines. Likewise, the consideration can be extended to other technical problem classes with corresponding datasets, such as bearings and battery charging cycles (Eker et al. 2012), in order to generate more significant evaluation results for the domain of maintenance. Moreover, the prognostic decision support task can be replaced by other DSA problem classes such as diagnostic or prescriptive tasks. Finally, the domain itself is also interchangeable. To this end, some first examinations have already been carried out to apply the general procedure to entirely different settings and verify the transferability of results. These examinations cover, for example, the field of computer vision and, more specifically, the problem class of 3D object detection (Friederich and Zschech 2020) as well as the field of business process management with a focus on next-step prediction (Heinrich et al. 2020). In subsequent research, the findings from these transfer studies will be used to improve the initial method proposal and provide a stronger formalization for better applicability.

Method Application for Solution Development

Finally, the applicability of the last artifact is exclusively limited to the domain of industrial maintenance and, more particularly, the problem class of prognostic maintenance when facing missing label information. Thus, by taking into account the concrete circumstances of a milling scenario at a German car manufacturer, a novel solution approach was developed to overcome the situation of inefficient maintenance strategy. Nevertheless, the conceptualization of the solution approach, as well as the technical realization, were outlined at a sufficiently generic level. In this way, the proposed solution can be transferred to similar problem classes where machine tools are subject to continuous wear and tear. Thus, instead of focusing on milling scenarios, the application could be extended to other settings such as those involving cutting, grinding, drilling, polishing, or similar operations, since only data collections were used that were expected to be recorded by default in industry.

On the downside, however, the proposed solution approach still lacks an in-depth evaluation as it could not yet be tested in real process executions. Although the findings were discussed with responsible machine operators in each development step, the overall approach has not yet been fully applied under proper conditions. Thus, future research should consider both a generalization of the solution approach by considering additional settings as well as a verification at the case level to evaluate the approach's feasibility in operational use.

7 Concluding Remarks

Over the last decades, efforts in research and practice have resulted in a variety of approaches that aim to facilitate the implementation of DSA projects so that broader user groups can conduct them more independently instead of permanently relying on fully equipped DSA professionals. For example, structured procedure models offer stepwise instructions for all relevant phases; DSA software solutions provide standardized functionalities, and intelligent assistants guide users in specific tasks such as choosing suitable analysis operators and parameters. All these approaches have in common that they encapsulate required multidisciplinary knowledge and codify best practices in the form of tools and methods in order to be reusable for a large group of users.

Nevertheless, the crucial challenge remains to reconcile the specificity of a domain with the possibilities of data-driven, analytical methods. To this end, this thesis provided several complementary artifacts to bridge the gap between the DSA world and the specific circumstances of a data-intensive domain. Each proposed artifact contributes differently to this goal. For example, (i) the systematization framework serves as a tool for communication purposes between different stakeholders such as domain experts and modeling specialists; (ii) the text-based IAS supports novice users to select a suitable class of analysis methods while expressing their problem space with domain-specific terms; and (iii) the taxonomic evaluation framework reveals which data-driven methods are adequate under certain domain-related conditions.

The particular domain focus of this thesis was primarily on the field of industrial maintenance. Nevertheless, an attempt was made to keep the artifacts sufficiently generic in order to ensure a high level of general applicability. Likewise, the achieved contributions show a high degree of novelty as there are currently only a limited number of initiatives dealing with similar research topics. On this note, possible connections to adjacent initiatives have been highlighted where the author of this thesis expects valuable synergies.

On the downside, it has to be acknowledged that some results have not yet reached full maturity. First and foremost, this requires verification of several artifacts in other contexts and conducting additional evaluation studies under more realistic conditions. Therefore, a critical reflection on open issues has been carried out in the previous chapter. However, despite facing several limitations, the derived findings are not less valuable. Instead, they fruitfully complement the field by providing new stimuli for other researchers and practitioners and by constituting the basis for subsequent work. The corresponding prospects for future efforts have been outlined in a detailed research agenda, which the author of this thesis is willing to tackle in upcoming research projects together with inspired collaborators.

Bibliography

Abbasi, A., Sarker, S., and Chiang, R. 2016. "Big Data Research in Information Systems: Toward an Inclusive Research Agenda," *Journal of the Association for Information Systems* (17:2), i–xxxii. (https://doi.org/10.17705/1jais.00423).

Accorsi, R., Manzini, R., Pascarella, P., Patella, M., and Sassi, S. 2017. "Data Mining and Machine Learning for Condition-Based Maintenance," *Procedia Manufacturing* (11), pp. 1153–1161. (https://doi.org/10.1016/j.promfg.2017.07.239).

Agarwal, R., and Dhar, V. 2014. "Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research," *Information Systems Research* (25:3), pp. 443–448. (https://doi.org/10.1287/isre.2014.0546).

Aggarwal, C. C., and Zhai, C. (eds.). 2012. *Mining Text Data*, Boston, MA: Springer US. (https://doi.org/10.1007/978-1-4614-3223-4).

Ahmad, R., and Kamaruddin, S. 2012. "An Overview of Time-Based and Condition-Based Maintenance in Industrial Application," *Computers & Industrial Engineering* (63:1), pp. 135–149. (https://doi.org/10.1016/j.cie.2012.02.002).

Ahmadzadeh, F., and Lundberg, J. 2014. "Remaining Useful Life Estimation: Review," *International Journal of System Assurance Engineering and Management* (5:4), pp. 461–474. (https://doi.org/10.1007/s13198-013-0195-0).

An, D., Kim, N. H., and Choi, J.-H. 2015. "Practical Options for Selecting Data-Driven or Physics-Based Prognostics Algorithms with Reviews," *Reliability Engineering & System Safety* (133), pp. 223–236. (https://doi.org/10.1016/j.ress.2014.09.014).

Ao, Y. H. 2011. "A Review on Development and Trend of Intelligent Maintenance System," *Advanced Materials Research* (314–316), pp. 2365–2369. (https://doi.org/10.4028/www.scientific.net/AMR.314-316.2365).

Ayankoya, K., Calitz, A., and Greyling, J. 2014. "Intrinsic Relations between Data Science, Big Data, Business Analytics and Datafication," in *Proceedings of the Southern African Institute for Computer Scientist and Information Technologists Annual Conference*, Centurion, South Africa: ACM Press, pp. 192–198. (https://doi.org/10.1145/2664591.2664619).

Babu, G. S., Zhao, P., and Li, X. L. 2016. "Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life," in *Database Systems for Advanced Applications*, Lecture Notes in Computer Science, Springer, Cham, pp. 214–228. (https://doi.org/10.1007/978-3-319-32025-0_14).

Bley, K., Leyh, C., and Schäffer, T. 2016. "Digitization of German Enterprises in the Production Sector - Do They Know How 'Digitized' They Are?," in *Proceedings of the 22nd Americas Conference on Information Systems (AMCIS)*, San Diego, CA, USA. (http://aisel.aisnet.org/amcis2016/EntSys/Presentations/9).

Boell, S. K., and Cecez-Kecmanovic, D. 2014. "A Hermeneutic Approach for Conducting Literature Reviews and Literature Searches," *Communications of the Association for Information Systems* (34:12), pp. 257–286. (https://doi.org/10.17705/1CAIS.03412).

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. 2017. "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics* (5), pp. 135–146. (https://doi.org/10.1162/tacl_a_00051).

Bousdekis, A., Magoutas, B., Apostolou, D., and Mentzas, G. 2018. "Review, Analysis and Synthesis of Prognostic-Based Decision Support Methods for Condition Based Maintenance," *Journal of Intelligent Manufacturing* (29:6), pp. 1303–1316. (https://doi.org/10.1007/s10845-015-1179-5).

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., and Cleven, A. 2009. “Reconstructing the Giant: On the Importance of Rigour in Documenting The Literature Search Process,” in *Proceedings of the 17th European Conference on Information Systems (ECIS)*, Verona, Italy.

Brodsky, A., Shao, G., Krishnamoorthy, M., Narayanan, A., Menasce, D., and Ak, R. 2015. “Analysis and Optimization in Smart Manufacturing Based on a Reusable Knowledge Base for Process Performance Models,” in *IEEE International Conference on Big Data*, Santa Clara, CA, USA: IEEE, pp. 1418–1427. (https://doi.org/10.1109/BigData.2015.7363902).

Buschmann, F. (ed.). 1996. *Pattern-Oriented Software Architecture: A System of Patterns*, Chichester, New York: Wiley.

Chen, H., Chiang, R. H. L., and Storey, V. C. 2012. “Business Intelligence and Analytics: From Big Data to Big Impact,” *MIS Quarterly* (36:4), pp. 1165–1188. (https://doi.org/10.2307/41703503).

Chollet, F. 2018. *Deep Learning with Python*, Shelter Island, New York: Manning Publications Co.

Coble, J. B. 2010. “Merging Data Sources to Predict Remaining Useful Life – An Automated Method to Identify Prognostic Parameters,” Knoxville: University of Tennessee.

Constantiou, I. D., and Kallinikos, J. 2015. “New Games, New Rules: Big Data and the Changing Context of Strategy:,” *Journal of Information Technology*, SAGE PublicationsSage UK: London, England. (https://doi.org/10.1057/jit.2014.17).

Dabab, M., Freiling, M., Rahman, N., and Sagalowicz, D. 2018. “A Decision Model for Data Mining Techniques,” in *Portland International Conference on Management of Engineering and Technology (PICMET)*, Honolulu, HI, USA: IEEE, pp. 1–8. (https://doi.org/10.23919/PICMET.2018.8481953).

Danubianu, M. 2008. “Design of an Expert System for Efficient Selection of Data Mining Method,” Universitatea Tehnică Gheorghe Asachi, p. 10.

Dragomir, O. E., Gouriveau, R., Dragomir, F., Minca, E., and Zerhouni, N. 2009. “Review of Prognostic Problem in Condition-Based Maintenance,” in *European Control Conference (ECC)*, Budapest: IEEE, pp. 1587–1592. (https://doi.org/10.23919/ECC.2009.7074633).

Eckert, S., and Ehmke, J. F. 2017. “Classification of Data Analysis Tasks for Production Environments,” in *Business Information Systems Workshops* (Vol. 263), W. Abramowicz, R. Alt, and B. Franczyk (eds.), Cham: Springer International Publishing, pp. 399–407. (https://doi.org/10.1007/978-3-319-52464-1_37).

Eker, O. F., Camci, F., and Jennions, I. K. 2012. “Major Challenges in Prognostics: Study on Benchmarking Prognostics Datasets,” in *European Conference of the Prognostics and Health Management Society*, Dresden, Germany, pp. 148–155.

Elattar, H. M., Elminir, H. K., and Riad, A. M. 2016. “Prognostics: A Literature Review,” *Complex & Intelligent Systems* (2:2), pp. 125–154. (https://doi.org/10.1007/s40747-016-0019-3).

Everitt, B. S., Landau, S., Leese, M., and Stahl, D. 2011. *Cluster Analysis*, (5th ed.), Wiley Series in Probability and Statistics, Chichester, West Sussex, UK: Wiley.

Famurewa, S. M., Zhang, L., and Asplund, M. 2017. “Maintenance Analytics for Railway Infrastructure Decision Support,” *Journal of Quality in Maintenance Engineering* (23:3), pp. 310–325. (https://doi.org/10.1108/JQME-11-2016-0059).

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. 1996. “From Data Mining to Knowledge Discovery in Databases,” *AI Magazine* (17:3), pp. 37–54. (https://doi.org/10.1609/aimag.v17i3.1230).

Fichman, R. G., Dos Santos, B. L., and Zheng, Z. E. 2014. “Digital Innovation as a Fundamental and Powerful Concept in the Information Systems Curriculum,” *MIS Quarterly* (38:2), pp. 329–343. (https://doi.org/10.25300/MISQ/2014/38.2.01).

Flath, C. M., and Stein, N. 2018. “Towards a Data Science Toolbox for Industrial Analytics Applications,” *Computers in Industry* (94), pp. 16–25. (https://doi.org/10.1016/j.compind.2017.09.003).

Friederich, J., and Zschech, P. 2020. “Review and Systematization of Solutions for 3D Object Detection,” in *Proceedings of the 15th International Conference on Wirtschaftsinformatik (WI)*, Potsdam, Germany: GITO Verlag, pp. 1699–1711. (https://doi.org/10.30844/wi_2020_r2-friedrich).

Gouriveau, R., Ramasso, E., and Zerhouni, N. 2013. “Strategies to Face Imbalanced and Unlabelled Data in PHM Applications,” *Chemical Engineering Transactions* (33), pp. 115–120. (https://doi.org/10.3303/CET1333020).

Goyal, D., and Pabla, B. S. 2015. “Condition Based Maintenance of Machine Tools - A Review,” *CIRP Journal of Manufacturing Science and Technology* (10), pp. 24–35. (https://doi.org/10.1016/j.cirpj.2015.05.004).

Gregor, S. 2006. “The Nature of Theory in Information Systems,” *MIS Quarterly* (30:3), pp. 611–642. (https://doi.org/10.2307/25148742).

Gregor, S., and Hevner, A. R. 2013. “Positioning and Presenting Design Science Research for Maximum Impact,” *MIS Quarterly* (37:2), pp. 337–355. (https://doi.org/10.25300/MISQ/2013/37.2.01).

Gutschi, C., Furian, N., Suschnigg, J., Neubacher, D., and Voessner, S. 2019. “Log-Based Predictive Maintenance in Discrete Parts Manufacturing,” *Procedia CIRP* (79), pp. 528–533. (https://doi.org/10.1016/j.procir.2019.02.098).

Hashemian, H. M., and Bean, W. C. 2011. “State-of-the-Art Predictive Maintenance Techniques*,” *IEEE Transactions on Instrumentation and Measurement* (60:10), pp. 3480–3492. (https://doi.org/10.1109/TIM.2009.2036347).

Heinrich, K., Zschech, P., Janiesch, C., and Bonin, M. 2020. “Ein Vergleich Aktueller Deep-Learning-Architekturen Zur Prognose von Prozessverhalten,” in *Proceedings of the 15th International Conference on Wirtschaftsinformatik (WI)*, Potsdam, Germany: GITO Verlag, pp. 876–892. (https://doi.org/10.30844/wi_2020_i1-heinrich).

Heng, A., Zhang, S., Tan, A. C. C., and Mathew, J. 2009. “Rotating Machinery Prognostics: State of the Art, Challenges and Opportunities,” *Mechanical Systems and Signal Processing* (23:3), pp. 724–739. (https://doi.org/10.1016/j.ymssp.2008.06.009).

Hesenius, M., Schwenzfeier, N., Meyer, O., Koop, W., and Gruhn, V. 2019. “Towards a Software Engineering Process for Developing Data-Driven Applications,” in *Proceedings of the 7th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering*, Montreal, Quebec, Canada: IEEE Press, pp. 35–41. (https://doi.org/10.1109/RAISE.2019.00014).

Hogl, O. M. J. 2003. “Eine wissensbasierte Benutzerschnittstelle für das Invisible Data Mining,” Dissertation, Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg. (https://opus4.kobv.de/opus4-fau/frontdoor/index/index/docId/55).

Horn, R., and Zschech, P. 2019. “Application of Process Mining Techniques to Support Maintenance-Related Objectives,” in *Proceedings of the 14th International Conference on Wirtschaftsinformatik (WI)*. (https://aisel.aisnet.org/wi2019/specialtrack01/papers/6).

Huber, S., Wiemer, H., Schneider, D., and Ihlenfeldt, S. 2019. “DMME: Data Mining Methodology for Engineering Applications – a Holistic Extension to the CRISP-DM Model,” *Procedia CIRP* (79), pp. 403–408. (https://doi.org/10.1016/j.procir.2019.02.106).

Iyyer, M., Manjunatha, V., Boyd-Graber, J., and Daumé III, H. 2015. "Deep Unordered Composition Rivals Syntactic Methods for Text Classification," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China: Association for Computational Linguistics, pp. 1681–1691. (https://doi.org/10.3115/v1/P15-1162).

Jardine, A. K. S., Lin, D., and Banjevic, D. 2006. "A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance," *Mechanical Systems and Signal Processing* (20:7), pp. 1483–1510. (https://doi.org/10.1016/j.ymssp.2005.09.012).

Javed, K., Gouriveau, R., and Zerhouni, N. 2017. "State of the Art and Taxonomy of Prognostics Approaches, Trends of Prognostics Applications and Open Issues towards Maturity at Different Technology Readiness Levels," *Mechanical Systems and Signal Processing* (94), pp. 214–236. (https://doi.org/10.1016/j.ymssp.2017.01.050).

Karim, R., Westerberg, J., Galar, D., and Kumar, U. 2016. "Maintenance Analytics – The New Know in Maintenance," *IFAC-PapersOnLine* (49:28), pp. 214–219. (https://doi.org/10.1016/j.ifacol.2016.11.037).

Kerschke, P., Hoos, H. H., Neumann, F., and Trautmann, H. 2019. "Automated Algorithm Selection: Survey and Perspectives," *Evolutionary Computation* (27:1), pp. 3–45. (https://doi.org/10.1162/evco_a_00242).

Khelif, R., Chebel-Morello, B., Malinowski, S., Laajili, E., Fnaiech, F., and Zerhouni, N. 2017. "Direct Remaining Useful Life Estimation Based on Support Vector Regression," *IEEE Transactions on Industrial Electronics* (64:3), pp. 2276–2285. (https://doi.org/10.1109/TIE.2016.2623260).

Kleindienst, D., Pfleger, R., and Schoch, M. 2015. "The Business Alignment of Social Media Analytics," in *Proceedings of the 23rd European Conference on Information Systems (ECIS)*, Muenster, Germany. (https://doi.org/10.18151/7217389).

Kothamasu, R., Huang, S. H., and VerDuin, W. H. 2006. "System Health Monitoring and Prognostics - A Review of Current Paradigms and Practices," *The International Journal of Advanced Manufacturing Technology* (28:9–10), pp. 1012–1024. (https://doi.org/10.1007/s00170-004-2131-6).

Kowsari, Jafari Meimandi, Heidarysafa, Mendu, Barnes, and Brown. 2019. "Text Classification Algorithms: A Survey," *Information* (10:4), p. 150. (https://doi.org/10.3390/info10040150).

LeCun, Y., Bengio, Y., and Hinton, G. 2015. "Deep Learning," *Nature* (521), pp. 436–444. (https://doi.org/10.1038/nature14539).

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., and Siegel, D. 2014. "Prognostics and Health Management Design for Rotary Machinery Systems - Reviews, Methodology and Applications," *Mechanical Systems and Signal Processing* (42:1–2), pp. 314–334. (https://doi.org/10.1016/j.ymssp.2013.06.004).

Lei, Y., Li, Naipeng, Guo, L., Li, Ningbo, Yan, T., and Lin, J. 2018. "Machinery Health Prognostics: A Systematic Review from Data Acquisition to RUL Prediction," *Mechanical Systems and Signal Processing* (104), pp. 799–834. (https://doi.org/10.1016/j.ymssp.2017.11.016).

Lemke, C., Budka, M., and Gabrys, B. 2015. "Metalearning: A Survey of Trends and Technologies," *Artificial Intelligence Review* (44:1), pp. 117–130. (https://doi.org/10.1007/s10462-013-9406-y).

Leturiondo, U., Salgado, O., Ciani, L., Galar, D., and Catelani, M. 2017. "Architecture for Hybrid Modelling and Its Application to Diagnosis and Prognosis with Missing Data," *Measurement* (108), pp. 152–162. (https://doi.org/10.1016/j.measurement.2017.02.003).

Li, N., Lei, Y., Yan, T., Li, N., and Han, T. 2018. "A Wiener Process Model-Based Method for Remaining Useful Life Prediction Considering Unit-to-Unit Variability," *IEEE Transactions on Industrial Electronics* (66:3), pp. 2092–2101. (https://doi.org/10.1109/TIE.2018.2838078).

Li, X., Ding, Q., and Sun, J.-Q. 2018. "Remaining Useful Life Estimation in Prognostics Using Deep Convolution Neural Networks," *Reliability Engineering & System Safety* (172), pp. 1–11. (https://doi.org/10.1016/j.ress.2017.11.021).

Lim, P., Goh, C. K., and Tan, K. C. 2016. "A Time Window Neural Network Based Framework for Remaining Useful Life Estimation," in *International Joint Conference on Neural Networks*, pp. 1746–1753. (https://doi.org/10.1109/IJCNN.2016.7727410).

Malhotra, P., TV, V., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., and Shroff, G. 2016. "Multi-Sensor Prognostics Using an Unsupervised Health Index Based on LSTM Encoder-Decoder," in *1st ACM SIGKDD Workshop on Machine Learning for Prognostics and Health Management*, San Francisco, CA, USA. (http://arxiv.org/abs/1608.06154).

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity | McKinsey," Technical Report, Technical Report, McKinsey Global Institute. (https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation).

Mariscal, G., Marbán, Ó., and Fernández, C. 2010. "A Survey of Data Mining and Knowledge Discovery Process Models and Methodologies," *The Knowledge Engineering Review* (25:2), pp. 137–166. (https://doi.org/10.1017/S0269888910000032).

Mathews, S. M. 2019. "Explainable Artificial Intelligence Applications in NLP, Biomedical, and Malware Classification: A Literature Review," in *Intelligent Computing* (Vol. 998), Advances in Intelligent Systems and Computing, K. Arai, R. Bhatia, and S. Kapoor (eds.), Cham: Springer International Publishing, pp. 1269–1292. (https://doi.org/10.1007/978-3-030-22868-2_90).

Meth, H., Mueller, B., and Maedche, A. 2015. "Designing a Requirement Mining System," *Journal of the Association for Information Systems* (16:9), pp. 799–837. (https://doi.org/10.17705/1jais.00408).

Morana, S., Maedche, A., Schacht, S., and Karlsruhe Institute of Technology. 2019. "Designing Process Guidance Systems," *Journal of the Association for Information Systems*, pp. 499–535. (https://doi.org/10.17705/1jais.00542).

Muchiri, P., Pintelon, L., Gelders, L., and Martin, H. 2011. "Development of Maintenance Function Performance Measurement Framework and Indicators," *International Journal of Production Economics* (131:1), pp. 295–302. (https://doi.org/10.1016/j.ijpe.2010.04.039).

Myers, M. D., and Newman, M. 2007. "The Qualitative Interview in IS Research: Examining the Craft," *Information and Organization* (17:1), pp. 2–26. (https://doi.org/10.1016/j.infoandorg.2006.11.001).

Nalchigar, S., and Yu, E. 2020. "Designing Business Analytics Solutions: A Model-Driven Approach," *Business & Information Systems Engineering* (62:1), pp. 61–75. (https://doi.org/10.1007/s12599-018-0555-z).

Nalchigar, S., Yu, E., Obeidi, Y., Carbajales, S., Green, J., and Chan, A. 2019. "Solution Patterns for Machine Learning," in *Advanced Information Systems Engineering* (Vol. 11483), Lecture Notes in Computer Science, P. Giorgini and B. Weber (eds.), Cham: Springer International Publishing, pp. 627–642. (https://doi.org/10.1007/978-3-030-21290-2_39).

Nickerson, R. C., Varshney, U., and Muntermann, J. 2013. "A Method for Taxonomy Development and Its Application in Information Systems," *European Journal of Information Systems* (22:3), pp. 336–359. (https://doi.org/10.1057/ejis.2012.26).

Pan, Y., Er, M. J., Li, X., Yu, H., and Gouriveau, R. 2014. “Machine Health Condition Prediction via Online Dynamic Fuzzy Neural Networks,” *Engineering Applications of Artificial Intelligence* (35), pp. 105–113. (https://doi.org/10.1016/j.engappai.2014.05.015).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. 2011. “Scikit-Learn: Machine Learning in Python,” *Journal of Machine Learning Research* (12:85), pp. 2825–2830.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. “A Design Science Research Methodology for Information Systems Research,” *Journal of Management Information Systems* (24:3), pp. 45–77. (https://doi.org/10.2753/MIS0742-1222240302).

Peng, Y., Dong, M., and Zuo, M. J. 2010. “Current Status of Machine Prognostics in Condition-Based Maintenance: A Review,” *The International Journal of Advanced Manufacturing Technology* (50:1–4), pp. 297–313. (https://doi.org/10.1007/s00170-009-2482-0).

Prajapati, A., Bechtel, J., and Ganesan, S. 2012. “Condition Based Maintenance: A Survey,” *Journal of Quality in Maintenance Engineering* (18:4), pp. 384–400. (https://doi.org/10.1108/13552511211281552).

Ramannavar, M., and Sidnal, N. S. 2016. “Big Data and Analytics—A Journey Through Basic Concepts to Research Issues,” in *Proceedings of the International Conference on Soft Computing Systems*, Advances in Intelligent Systems and Computing, L. P. Suresh and B. K. Panigrahi (eds.), New Delhi: Springer India, pp. 291–306. (https://doi.org/10.1007/978-81-322-2674-1_29).

Ramasso, E., and Saxena, A. 2014. “Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets,” *International Journal of Prognostics and Health Management* (5:2), pp. 1–15.

Ran, Y., Zhou, X., Lin, P., Wen, Y., and Deng, R. 2019. *A Survey of Predictive Maintenance: Systems, Purposes and Approaches*. (http://arxiv.org/abs/1912.07383, accessed April 25, 2020).

RapidMiner. 2020. “Lightning Fast Unified Data Science Platform | RapidMiner,” *RapidMiner*. (https://rapidminer.com/products/, accessed July 15, 2019).

Russo, B. 2016. “The Need for Data Analysis Patterns (in Software Engineering),” in *Perspectives on Data Science for Software Engineering*, Elsevier, pp. 19–23. (https://doi.org/10.1016/B978-0-12-804206-9.00004-0).

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., and Schwabacher, M. 2008. “Metrics for Evaluating Performance of Prognostic Techniques,” in *International Conference on Prognostics and Health Management*, Denver, USA, pp. 1–17. (https://doi.org/10.1109/PHM.2008.4711436).

Saxena, A., Goebel, K., Simon, D., and Eklund, N. 2008. “Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation,” in *International Conference on Prognostics and Health Management*, Denver, USA, pp. 1–9. (https://doi.org/10.1109/PHM.2008.4711414).

Schumann, C., Zschech, P., and Hilbert, A. 2016. “Das aufstrebende Berufsbild des Data Scientist: Vom Kompetenzwirrwarr zu spezifischen Anforderungsprofilen,” *HMD Praxis der Wirtschaftsinformatik* (53:4), pp. 453–466. (https://doi.org/10.1365/s40702-016-0214-0).

Schwabacher, M. 2005. “A Survey of Data-Driven Prognostics,” in *Infotech@Aerospace*, Arlington, Virginia: American Institute of Aeronautics and Astronautics. (https://doi.org/10.2514/6.2005-7002).

Seabold, S., and Perktold, J. 2010. “Statsmodels: Econometric and Statistical Modeling with Python,” in *Proceedings of the 9th Python in Science Conference*, Austin, Texas, pp. 92–96. (https://doi.org/10.25080/Majora-92bf1922-011).

Serban, F., Vanschoren, J., Kietz, J.-U., and Bernstein, A. 2013. “A Survey of Intelligent Assistants for Data Analysis,” *ACM Computing Surveys* (45:3), pp. 1–35. (https://doi.org/10.1145/2480741.2480748).

Si, X.-S., Wang, W., Hu, C.-H., and Zhou, D.-H. 2011. “Remaining Useful Life Estimation – A Review on the Statistical Data Driven Approaches,” *European Journal of Operational Research* (213:1), pp. 1–14. (https://doi.org/10.1016/j.ejor.2010.11.018).

Sipos, R., Fradkin, D., Moerchen, F., and Wang, Z. 2014. “Log-Based Predictive Maintenance,” in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, New York, USA: ACM Press, pp. 1867–1876. (https://doi.org/10.1145/2623330.2623340).

Sneath, P. H. A., and Sokal, R. R. 1973. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*, A Series of Books in Biology, San Francisco: W. H. Freeman.

Stefani, K., and Zschech, P. 2018. “Constituent Elements for Prescriptive Analytics Systems,” in *Proceedings of the 26th European Conference on Information Systems (ECIS)*, Portsmouth, UK. (https://aisel.aisnet.org/ecis2018_rp/39).

Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., and Beghi, A. 2015. “Machine Learning for Predictive Maintenance: A Multiple Classifier Approach,” *IEEE Transactions on Industrial Informatics* (11:3), pp. 812–820. (https://doi.org/10.1109/TII.2014.2349359).

Tao, M., Man, Z., Zheng, J., Cricenti, A., and Wang, W. 2016. “A New Dynamic Neural Modelling for Mechatronic System Prognostics,” in *International Conference on Advanced Mechatronic Systems*, pp. 437–442. (https://doi.org/10.1109/ICAMechS.2016.7813487).

Tsai, C.-W., Lai, C.-F., Chiang, M.-C., and Yang, L. T. 2014. “Data Mining for Internet of Things: A Survey,” *IEEE Communications Surveys & Tutorials* (16:1), pp. 77–97. (https://doi.org/10.1109/SURV.2013.103013.00206).

Vainshtein, R., Greenstein-Messica, A., Katz, G., Shapira, B., and Rokach, L. 2018. “A Hybrid Approach for Automatic Model Recommendation,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, Torino, Italy: ACM Press, pp. 1623–1626. (https://doi.org/10.1145/3269206.3269299).

Veldman, J., Wortmann, H., and Klingenberg, W. 2011. “Typology of Condition Based Maintenance,” *Journal of Quality in Maintenance Engineering* (17:2), pp. 183–202. (https://doi.org/10.1108/13552511111134600).

Vogl, G. W., Weiss, B. A., and Helu, M. 2019. “A Review of Diagnostic and Prognostic Capabilities and Best Practices for Manufacturing,” *Journal of Intelligent Manufacturing* (30:1), pp. 79–95. (https://doi.org/10.1007/s10845-016-1228-8).

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., and Gnanzou, D. 2015. “How ‘Big Data’ Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study,” *International Journal of Production Economics* (165), pp. 234–246. (https://doi.org/10.1016/j.ijpe.2014.12.031).

Wang, T. 2010. “Trajectory Similarity Based Prediction for Remaining Useful Life Estimation,” Doctoral Dissertation, Doctoral Dissertation, University of Cincinnati.

Wang, Z., Tang, W., and Pi, D. 2017. “Trajectory Similarity-Based Prediction with Information Fusion for Remaining Useful Life,” in *Intelligent Data Engineering and Automated Learning*, Lecture Notes in Computer Science, Springer, Cham, pp. 270–278. (https://doi.org/10.1007/978-3-319-68935-7_30).

Webster, J., and Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly* (26:2), xiii–xxiii.

Wilcox, R. R. 1989. "Adjusting for Unequal Variances When Comparing Means in One-Way and Two-Way Fixed Effects ANOVA Models," *Journal of Educational Statistics* (14:3), pp. 269–278. (https://doi.org/10.3102/10769986014003269).

Williams, R. J. 1995. "Adaptive State Representation and Estimation Using Recurrent Connectionist Networks," in *Neural Networks for Control*, W. T. Miller, R. S. Sutton, and P. J. Werbos (eds.), Cambridge: MIT Press, pp. 97–114. (https://ieeexplore.ieee.org/document/6300640).

Wirth, R., and Hipp, J. 2000. "CRISP-DM: Towards a Standard Process Model for Data Mining," in *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, pp. 29–39.

Zheng, S., Ristovski, K., Farahat, A., and Gupta, C. 2017. "Long Short-Term Memory Network for Remaining Useful Life Estimation," in *IEEE International Conference on Prognostics and Health Management*, pp. 88–95. (https://doi.org/10.1109/ICPHM.2017.7998311).

Zhu, L., Jiang, B., and Cheng, Y. 2016. "Life Prediction Methods Based on Data-Driven: Review and Trend," in *IEEE Chinese Guidance, Navigation and Control Conference*, Nanjing: IEEE, pp. 1682–1686. (https://doi.org/10.1109/CGNCC.2016.7829044).

Zschech, P. 2018. "A Taxonomy of Recurring Data Analysis Problems in Maintenance Analytics," in *Proceedings of the 26th European Conference on Information Systems (ECIS)*, Portsmouth, UK. (https://aisel.aisnet.org/ecis2018_rp/197).

Zschech, P., Bernien, J., and Heinrich, K. 2019. "Towards a Taxonomic Benchmarking Framework for Predictive Maintenance: The Case of NASA's Turbofan Degradation," in *Proceedings of the 40th International Conference on Information Systems (ICIS)*, Munich, Germany. (https://aisel.aisnet.org/icis2019/data_science/data_science/4).

Zschech, P., Fleißner, V., Baumgärtel, N., and Hilbert, A. 2018. "Data Science Skills and Enabling Enterprise Systems: Eine Erhebung von Kompetenzanforderungen und Weiterbildungsangeboten," *HMD Praxis der Wirtschaftsinformatik* (55:1), pp. 163–181. (https://doi.org/10.1365/s40702-017-0376-4).

Zschech, P., Heinrich, K., Bink, R., and Neufeld, J. S. 2019. "Prognostic Model Development with Missing Labels: A Condition-Based Maintenance Approach Using Machine Learning," *Business & Information Systems Engineering* (61:3), pp. 327–343. (https://doi.org/10.1007/s12599-019-00596-1).

Zschech, P., Heinrich, K., Horn, R., and Höschele, D. 2019. "Towards a Text-Based Recommender System for Data Mining Method Selection," in *Proceedings of the 25th Americas Conference on Information Systems (AMCIS)*, Cancún, Mexico. (https://aisel.aisnet.org/amcis2019/ai_semantic_for_intelligent_info_systems/ai_semantic_for_intelligent_info_systems/4).

Zschech, P., Heinrich, K., Pfitzner, M., and Hilbert, A. 2017. "Are You Up for the Challenge? Towards the Development of a Big Data Capability Assessment Model," in *Proceedings of the 25th European Conference on Information Systems (ECIS)*, Guimarães, Portugal, pp. 2613–2624. (https://aisel.aisnet.org/ecis2017_rip/14).

Zschech, P., Horn, R., Höschele, D., Janiesch, C., and Heinrich, K. 2020. "Intelligent User Assistance for Automated Data Mining Method Selection," *Business & Information Systems Engineering* (62:3), pp. 227–247. (https://doi.org/10.1007/s12599-020-00642-3).

Appendix I: Implementation Details

In this appendix, further information is provided on the technical realization of the implemented approaches in Chapter 4.3. The choice of the methods and parameters is based on findings derived from the examined C-MAPSS studies as well as experiments using five-fold cross-validation to select the best performing approaches. Furthermore, for the development of all pipelines, the training samples of the C-MAPSS collection are used. By contrast, the evaluation is performed on out-of-sample data using the corresponding test samples (Ramasso and Saxena 2014).

Employing the min-max transformation (Tao et al. 2016), all sensor measurements are transferred into a value range [0,1]. For datasets with multiple operational conditions (i.e., FD002, FD004), the rescaling approach is performed separately for each cluster of operational conditions. The identification of the clusters is performed using a k-means approach from the Python module *sklearn.cluster* (Pedregosa et al. 2011).

For the implementation of the locally weighted scatterplot smoothing as a variant of polynomial smoothing (Khelif et al. 2017), the *lowess* function of the Python library *statsmodels* (Seabold and Perktold 2010) is used. For the calculation of the parameter f as the quotient of the time window and the total length of a unit, the window size is set to 15 (N. Li et al. 2018).

To realize the filter approach for feature selection, the metrics *monotonicity*, *prognosability* and *trendability* are calculated for each feature (Coble 2010). Subsequently, all features are ranked based on an equally weighted score (w = 1/3), whereby the best eight are selected; the number eight is determined by conducting several cross-validation experiments in which the performance considerably drops using fewer features.

The two deep neural networks are implemented using the Python library *Keras* in combination with *TensorFlow* as a backend (Chollet 2018). The architecture of the LSTM network is adapted with slight modifications from Zheng et al. (2017), while the CNN is reconstructed following the example of Babu et al. (2016). The implemented architectures and the chosen parameters are summarized in Table 13 and Table 14. For compiling both networks, the *Adadelta optimizer* is chosen, and *callback early stopping* is used to terminate the training if the validation loss does not improve over several epochs. The *mean squared error* specifies the loss function. Additionally, following X. Li et al. (2018), a maximum value for RUL estimates is set to 125, and different time window sizes are chosen for each dataset ($tw_{FD001} = 30$, $tw_{FD002} = 20$, $tw_{FD003} = 30$, $tw_{FD004} = 15$).

For the implementation of the two similarity-based approaches, the first step of HI construction is performed using linear regression and a binary objective function. For this purpose, sensor measurements covering the first 10% of a cycle are assigned to 1, whereas the last 10% are assigned to 0. The regression parameters are then determined exclusively in this sample (Khelif et al. 2017). After that, a curve fitting approach is performed for *Similarity1,* following the examples of Wang (2010) and Wang et al. (2017). To this end, a second-order polynomial is used whose parameters are determined using least square fitting via the Python function *numpy.polyfit*. To assess the similarity between training and test units, an *information fusion* approach is implemented (Wang et al. 2017). In equivalence to the direct methods, a maximum value for the RUL estimates is set to 125. For *Similarity2,* a different approach for similarity matching is used, following the example of Malhotra et al. (2016). Instead of fitting a curve function, the HI created with the linear regression is smoothed. For this purpose, the *lowess* method described above is used again, whereby the time window is set to 15. Table 15 provides an overview of the implementations for both similarity-based approaches.

Layer ID	*Layer*	*Parameters*
1	LSTM	units = 64, return_sequences = true
	Dropout	rate = 0.2
2	LSTM	units = 64, return_sequences = true
	Dropout	rate = 0.2
3	LSTM	units = 8
	Dropout	rate = 0.2
4	LSTM	units = 8
	Dropout	rate = 0.2
5	Dense	units = 1

Table 13: Summary of the implemented LSTM architecture (adapted from Zheng et al. 2017)

Layer ID	*Layer*	*Parameters*
1	2D-Convolution	filters = 8, kernel_size = (features_length, 4), activation = ‘relu’
2	2D-Average Pooling	pool_size = (1, 2), strides = 2
3	2D-Convolution	filters = 14, kernel_size = (1, 3), activation = ’relu’
4	2D-Average Pooling	pool_size = (1, 2), strides = 2
5	Dense	layer_size = 1

Table 14: Summary of the implemented CNN architecture (adapted from Babu et al. 2016)

Parameter	*Similarity1*	*Similarity2*
HI construction	Linear regression	Linear regression
Objective function	Binary	Binary
Curve matching	Second-order polynomial	None
Time lag	Considered	Considered
Similarity score	Information fusion	Euclidian distance
Range of RUL	[0, 125]	[0, 125]

Table 15: Summary of the implemented similarity-based approaches

Appendix II: List of Publications

(* element of this thesis)

Bink, R., and Zschech, P. 2018. “Predictive Maintenance in der industriellen Praxis: Entwicklung eines Prognoseansatzes unter eingeschränkter Informationslage,” *HMD Praxis der Wirtschaftsinformatik* (55:3), pp. 552–565.

Friederich, J., and Zschech, P. 2020. “Review and Systematization of Solutions for 3D Object Detection,” in *Proceedings of the 15th International Conference on Wirtschaftsinformatik (WI)*, Potsdam, Germany: GITO Verlag, pp. 1699–1711.

Heinrich, K., Graf, J., Chen, J., Laurisch, J., and Zschech, P. 2020. “Fool Me Once, Shame On You, Fool Me Twice, Shame On Me: A Taxonomy of Attack and Defense Patterns for AI Security,” in *Proceedings of the 28th European Conference on Information Systems (ECIS)*, Marrakesh, Morocco.

Heinrich, K., Möller, B., Janiesch, C., and Zschech, P. 2019. “Is Bigger Always Better? Lessons Learnt from the Evolution of Deep Learning Architectures for Image Classification,” in *Proceedings of the 2019 Pre-ICIS SIGDSA Symposium*, Munich, Germany.

Heinrich, K., Roth, A., and Zschech, P. 2019. “Everything Counts: A Taxonomy of Deep Learning Approaches for Object Counting,” in *Proceedings of the 27th European Conference on Information Systems (ECIS)*, Stockholm-Uppsala, Sweden.

Heinrich, K., Zschech, P., Janiesch, C., and Bonin, M. 2020. “Ein Vergleich Aktueller Deep-Learning-Architekturen Zur Prognose von Prozessverhalten,” in *Proceedings of the 15th International Conference on Wirtschaftsinformatik (WI)*, Potsdam, Germany: GITO Verlag, pp. 876–892.

Heinrich, K., Zschech, P., Möller, B., Breithaupt, L., and Maresch, J. 2019. “Objekterkennung im Weinanbau – Eine Fallstudie zur Unterstützung von Winzertätigkeiten mithilfe von Deep Learning,” *HMD Praxis der Wirtschaftsinformatik* (56:5), pp. 964–985.

Heinrich, K., Zschech, P., Skouti, T., Griebenow, J., and Riechert, S. 2019. “Demystifying the Black Box: A Classification Scheme for Interpretation and Visualization of Deep Intelligent Systems,” in *Proceedings of the 25th Americas Conference on Information Systems (AMCIS)*, Cancún, Mexico.

Hilbert, A., and Zschech, P. 2016. “Process Analytics,” *WISU - Das Wirtschaftsstudium* (45:8/9), pp. 942–948.

Horn, R., and Zschech, P. 2019. “Application of Process Mining Techniques to Support Maintenance-Related Objectives,” in *Proceedings of the 14th International Conference on Wirtschaftsinformatik (WI)*, Siegen, Germany.

Könning, M., Heinrich, K., Zschech, P., and Leyh, C. 2018. “Analyzing Influences on Pivotal ITO Contract Features: A Quantitative Multi-Study Design with Evidence from Western Europe,” in *Proceedings of the 24th Americas Conference on Information Systems (AMCIS)*, Boston, USA.

Kruse, P., Kummer, C., and Zschech, P. 2014. “Existieren Wissensmanagement-Schulen? Eine Clusteranalyse von Wissensmanagement-Beiträgen aus den letzten 10 Jahren,” in *Online Communities: Technologies and Analyses for Networks in Industry, Research and Education: 17. Workshop GeNeMe ’14 Gemeinschaften in Neuen Medien*, Dresden, Germany.

Schumann, C., Zschech, P., and Hilbert, A. 2016. “Das aufstrebende Berufsbild des Data Scientist: Vom Kompetenzwirrwarr zu spezifischen Anforderungsprofilen,” *HMD Praxis der Wirtschaftsinformatik* (53:4), pp. 453–466.

Stefani, K., and Zschech, P. 2018. “Constituent Elements for Prescriptive Analytics Systems,” in *Proceedings of the 26th European Conference on Information Systems (ECIS)*, Portsmouth, UK.

*Zschech, P. 2018. “A Taxonomy of Recurring Data Analysis Problems in Maintenance Analytics,” in *Proceedings of the 26th European Conference on Information Systems (ECIS)*, Portsmouth, UK.

*Zschech, P., Bernien, J., and Heinrich, K. 2019. “Towards a Taxonomic Benchmarking Framework for Predictive Maintenance: The Case of NASA’s Turbofan Degradation,” in *Proceedings of the 40th International Conference on Information Systems (ICIS)*, Munich, Germany.

Zschech, P., Fleißner, V., Baumgärtel, N., and Hilbert, A. 2018. “Data Science Skills and Enabling Enterprise Systems: Eine Erhebung von Kompetenzanforderungen und Weiterbildungsangeboten,” *HMD Praxis der Wirtschaftsinformatik* (55:1), pp. 163–181.

*Zschech, P., Heinrich, K., Bink, R., and Neufeld, J. S. 2019. “Prognostic Model Development with Missing Labels: A Condition-Based Maintenance Approach Using Machine Learning,” *Business & Information Systems Engineering* (61:3), pp. 327–343.

*Zschech, P., Heinrich, K., Horn, R., and Höschele, D. 2019. “Towards a Text-Based Recommender System for Data Mining Method Selection,” in *Proceedings of the 25th Americas Conference on Information Systems (AMCIS)*, Cancún, Mexico.

Zschech, P., Heinrich, K., Pfitzner, M., and Hilbert, A. 2017. “Are You Up for the Challenge? Towards the Development of a Big Data Capability Assessment Model,” in *Proceedings of the 25th European Conference on Information Systems (ECIS)*, Guimarães, Portugal, pp. 2613–2624.

*Zschech, P., Horn, R., Höschele, D., Janiesch, C., and Heinrich, K. 2020. “Intelligent User Assistance for Automated Data Mining Method Selection,” *Business & Information Systems Engineering* (62:3), pp. 227–247.

Zschech, P., Pfitzner, M., and Hilbert, A. 2017. “Vom Controller zum Prozessanalysten,” *Controlling & Management Review* (61:4), pp. 24–33.

A Publication P1: Focus Area Systematization

Title	A Taxonomy of Recurring Data Analysis Problems in Maintenance Analytics
Author(s)	Patrick Zschech (PZ)
Publication	Zschech, P. 2018. “A Taxonomy of Recurring Data Analysis Problems in Maintenance Analytics,” in *Proceedings of the 26th European Conference on Information Systems (ECIS)*, Portsmouth, UK.
Link	`https://aisel.aisnet.org/ecis2018_rp/197/`
Ranking	VBH-Jourqual 3: B WKWI 2008: A
Abstract	Modern maintenance strategies increasingly focus on vast amounts of diverse data and multi-faceted analytical approaches in order to make efficient use of given resources and unveil hidden potentials. While there is often no universal solution approach to a specific case at hand, it is still possible to observe recurring problem classes for which generic solution templates can be applied and thus the establishment of a reusable knowledge base appears beneficial. To this end, we apply a taxonomy development approach to identify and systematize dimensions and characteristics of recurring data analysis problems in data-driven maintenance scenarios. Our research method integrates findings from a systematic literature review and expert interviews with data scientists from industry. Thus, we add descriptive theory to the field of maintenance analytics and propose a taxonomy that distinguishes between analytical maintenance objectives, data characteristics and analytical techniques.

Table 16: Summary of publication P1

B Publication P2: Focus Area Method Selection

Title	Towards a Text-based Recommender System for Data Mining Method Selection
Author(s)	Patrick Zschech (PZ), Kai Heinrich (KH), Richard Horn (RH), Daniel Höschele (DH)
Publication	Zschech, P., Heinrich, K., Horn, R., and Höschele, D. 2019. "Towards a Text-Based Recommender System for Data Mining Method Selection," in *Proceedings of the 25th Americas Conference on Information Systems (AMCIS)*, Cancún, Mexico.
Link	`https://aisel.aisnet.org/amcis2019/ai_semantic_for_intelligent_info_systems/ai_semantic_for_intelligent_info_systems/4/`
Ranking	VBH-Jourqual 3: D WKWI 2008: B
Abstract	The task of mapping a domain-specific problem space to an adequate set of data mining (DM) methods is a crucial step in data science projects. While there have been several efforts for automated method selection in general, only few approaches consider the particularities of problem contexts expressed in domain-specific language. Therefore, we propose the concept of a text-based recommender system (TBRS) which takes problem descriptions articulated in domain language as inputs and then recommends the best suitable class of DM methods. Following a design science research methodology, the current focus is on the initial steps of motivating the problem and conducting a requirements analysis. In particular, we outline the problem setting using an exemplary scenario from industrial practice and derive requirements towards an adequate solution artifact. Subsequently, we discuss potential TBRS methods with regard to requirement fulfillment while organizing both methods and requirements in a structured framework. Finally, we conclude the paper, discuss limitations and draw an outlook.

Table 17: Summary of publication P2

C Publication P3: Focus Area Method Selection

Title	Intelligent User Assistance for Automated Data Mining Method Selection
Author(s)	Patrick Zschech (PZ), Richard Horn (RH), Daniel Höschele (DH), Christian Janiesch (CJ), Kai Heinrich (KH)
Publication	Zschech, P., Horn, R., Höschele, D., Janiesch, C., and Heinrich, K. 2020. "Intelligent User Assistance for Automated Data Mining Method Selection," *Business & Information Systems Engineering* (62:3), pp. 227–247.
Link	`https://doi.org/10.1007/s12599-020-00642-3`
Ranking	VBH-Jourqual 3: B WKWI 2008: A
Abstract	In any data science and analytics project, the task of mapping a domain-specific problem to an adequate set of data mining methods by experts of the field is a crucial step. However, these experts are not always available and data mining novices may be required to perform the task. While there are several research efforts for automated method selection as a means of support, only a few approaches consider the particularities of problems expressed in the natural and domain-specific language of the novice. The study proposes the design of an intelligent assistance system that takes problem descriptions articulated in natural language as an input and offers advice regarding the most suitable class of data mining methods. Following a design science research approach, the paper (i) outlines the problem setting with an exemplary scenario from industrial practice, (ii) derives design requirements, (iii) develops design principles and proposes design features, (iv) develops and implements the IT artifact using several methods such as embeddings, keyword extractions, topic models, and text classifiers, (v) demonstrates and evaluates the implemented prototype based on different classification pipelines, and (vi) discusses the results' practical and theoretical contributions. The best performing classification pipelines show high accuracies when applied to validation data and are capable of creating a suitable mapping that exceeds the performance of joint novice assessments and simpler means of text mining. The research provides a promising foundation for further enhancements, either as a stand-alone intelligent assistance system or as an add-on to already existing data science and analytics platforms.

Table 18: Summary of publication P3

D Publication P4: Focus Area Method Evaluation

Title	Towards a Taxonomic Benchmarking Framework for Predictive Maintenance: The Case of NASA's Turbofan Degradation
Author(s)	Patrick Zschech (PZ), Jonas Bernien (JB), Kai Heinrich (KH)
Publication	Zschech, P., Bernien, J., and Heinrich, K. 2019. "Towards a Taxonomic Benchmarking Framework for Predictive Maintenance: The Case of NASA's Turbofan Degradation," in *Proceedings of the 40th International Conference on Information Systems (ICIS)*, Munich, Germany.
Link	`https://aisel.aisnet.org/icis2019/data_science/data_science/4/`
Ranking	VBH-Jourqual 3: A WKWI 2008: A
Abstract	The availability of datasets for analytical solution development is a common bottleneck in data-driven predictive maintenance. Therefore, novel solutions are mostly based on synthetic benchmarking examples, such as NASA's C-MAPSS datasets, where researchers from various disciplines like artificial intelligence and statistics apply and test their methodical approaches. The majority of studies, however, only evaluate the overall solution against a final prediction score, where we argue that a more fine-grained consideration is required distinguishing between detailed method components to measure their particular impact along the prognostic development process. To address this issue, we first conduct a literature review resulting in more than one hundred studies using the C-MAPSS datasets. Subsequently, we apply a taxonomy approach to receive dimensions and characteristics that decompose complex analytical solutions into more manageable components. The result is a first draft of a systematic benchmarking framework as a more comparable basis for future development and evaluation purposes.

Table 19: Summary of publication P4

E Publication P5: Focus Area Method Application

Title	Prognostic Model Development with Missing Labels - A Condition-Based Maintenance Approach Using Machine Learning
Author(s)	Patrick Zschech (PZ), Kai Heinrich (KH), Raphael Bink (RB), Janis S. Neufeld (JN)
Publication	Zschech, P., Heinrich, K., Bink, R., and Neufeld, J. S. 2019. “Prognostic Model Development with Missing Labels: A Condition-Based Maintenance Approach Using Machine Learning,” *Business & Information Systems Engineering* (61:3), pp. 327–343.
Link	`https://doi.org/10.1007/s12599-019-00596-1`
Ranking	VBH-Jourqual 3: B WKWI 2008: A
Abstract	Condition-based maintenance (CBM) has emerged as a proactive strategy for determining the best time for maintenance activities. In this paper, a case of a milling process with imperfect maintenance at a German automotive manufacturer is considered. Its major challenge is that only data with missing labels are available, which does not provide a sufficient basis for classical prognostic maintenance models. To overcome this shortcoming, a data science study is carried out that combines several analytical methods, especially from the field of machine learning (ML). These include time-domain and time–frequency domain techniques for feature extraction, agglomerative hierarchical clustering and time series clustering for unsupervised pattern detection, as well as a recurrent neural network for prognostic model training. With the approach developed, it is possible to replace decisions that were made based on subjective criteria with data driven decisions to increase the tool life of the milling machines. The solution can be employed beyond the presented case to similar maintenance scenarios as the basis for decision support and prognostic model development. Moreover, it helps to further close the gap between ML research and the practical implementation of CBM.

Table 20: Summary of publication P5

Statutory Declaration

I herewith declare that I have composed the present thesis with the title

Data Science and Analytics in Industrial Maintenance:
Selection, Evaluation, and Application of Data-Driven Methods

independently and without the use of any sources or aids other than those indicated, that all parts and passages of the thesis that have been quoted either literally or by content from other sources are marked as such, and that the thesis has not yet been submitted in the same or similar form to any examination authority.

Dresden, June 8, 2020

Patrick Zschech