POTENTIALS, LIMITS AND CHALLENGES OF USING DATA SCIENCE METHODS TO IMPROVE QUALITY PROCESSES IN MANUFACTURING INDUSTRY

Jens Kiefer[1], Lisa Ollinger[1], Melina Precht[2] and Kai Beer[2]

[1]University of Applied Sciences Ulm
Prittwitzstraße 10, 89075 Ulm, Germany
[2]Heidelberg Manufacturing Deutschland GmbH
Egelsee 10, 73340 Amstetten, Germany

ABSTRACT

This paper presents potentials, limits and challenges of using data science methods to improve quality processes in manufacturing industry. Apart from scientific discussions, this paper focuses on the specific application of data science methods based on real conditions and data from manufacturing industry. The application data, processes and all technical and organizational conditions originate from the foundry sector. As a first step, all necessary terms are explained and the most important processes and characteristics of the considered application scenario 'quality processes of a contract manufacturer for cast components' are illustrated. By means of two practical examples, potentials, limits and challenges of using data science methods in operational practice are critically evaluated. In terms of a socio-technical overall view, not only methodical issues but also information-technical and organizational aspects are discussed. The collected findings ultimately lead to recommendations for action, which are intended to give companies a kind of framework for a meaningful economic use of data science methods in operational practice.

KEYWORDS

Data Science, Artificial Intelligence (AI), Quality Processes, Manufacturing Industry

1. INTRODUCTION AND RESEARCH QUESTION

According to several studies (e.g. EIT, 2020), the profitable use of artificial intelligence (AI) in manufacturing industry is of outstanding strategic importance for most national economies. In a study on behalf of the German Ministry for Economic Affairs and Energy (BMWi) (PAiCE, 2018), it is even predicted that AI will be jointly responsible for around a third of the average economic growth in the future. In this context, quality management, data analytics and automation of processes in manufacturing industry are classified as having a particularly high potential (IFS, 2018).

Between these identified AI potentials and the current implementation status in manufacturing companies is (still) a large gap, which can be often referred to the following three reasons: inadequate application-oriented AI research activities, a reluctant technology transfer from science to companies as well as deficits in the development and implementation of AI technologies with a cross-sectional character such as machine learning, action planning and optimization (PAiCE, 2018).

In cooperation with the company Heidelberg Manufacturing Deutschland GmbH (HMD), the University of Applied Sciences Ulm (THU) wants to take on this identified topic field in the context of quality and process management in manufacturing industry. The superordinate research question is: What potentials do data science and particularly AI techniques offer in production-related engineering and manufacturing processes with regard to quality and process management, how can these potentials be most effectively developed and what are limits, challenges and recommendations for action of using AI concepts in this topic field?

Focus of the cooperation between the THU and HMD is the coherent processing of the following two subject areas:

- Development of a standards-based, semantic quality data and quality service model that transparently provides all relevant quality-related development, planning, production and customer data (and functions) to the identified applications within the quality process chain
- Development and use of effective and efficient AI algorithms for continuous planning and control of quality processes along the entire product life cycle

The scientific goals and the associated innovations are the development of a standards-based (and therefore system-independent) quality data and quality service model, the development and provision of modularly structured and independently executable services in the form of a standardized module library as well as in the implementation of a consistent overall solution. First, this paper focuses on the illustration of potentials, limits and challenges of using data science methods to improve quality processes in manufacturing industry.

2. DATA SCIENCE, AI AND QUALITY PROCESSES

2.1 Need for Data Science in Production

During the last decade, the massive expansion of digitization dominated the technical progress within the production industry. In Germany, these developments fall under the umbrella of the 'Industrie 4.0' initiative that internationally gained in great significance (Wahlster, 2021).

Digitization enables the rise of smart sensors that deliver information about production processes and products in real-time. They deliver information represented as data via communication standards that support both the horizontal and vertical integration, like OPC-UA and MQTT (Shi et al, 2020). Thus, potential data sources and data sinks can be seamlessly interconnected within and outside of the production system. This situation leads to drastically increasing amounts of data that provide substantial benefit for the processes at the company (PAiCE, 2018). However, the sheer volume of data makes it literally impossible for a human to parse it in a reasonable time frame (Ozdemir and Sunil, 2018). Furthermore, the value of the stored data is zero unless it is acted upon (Kotu and Bala, 2018). This where data science comes into play to make use of data as a resource in order to gain new knowledge about the production process.

2.2 Data Science and Related Disciplines

Data science is the art and science of acquiring knowledge and new insights through data (Ozdemir and Sunil, 2018). It requires the combination of three disciplines: computer programming, mathematics and statistics, and domain knowledge (Grus, 2019). The mathematical methods constitute the base for the algorithms that find patterns and relationships within data. These algorithms are translated into automatically executable programs by means of computer science. The domain knowledge depends on the respective application field. It is indispensable to provide the required data, to supervise the algorithms and to interpret the results correctly depending on the context of the application domain.

Artificial intelligence (AI), machine learning (ML) and data science are all related to each other and are often used interchangeably in popular media (Kotu and Bala, 2018). AI gives machines the capability of mimicking human behavior, particularly cognitive functions like facial recognition and automated driving. ML can be classified either as sub-field or as one of the tools of AI. It deals with gaining the capability of learning from experience whereas the experience for machines comes in the form of data (Kotu and Bala, 2018). Those methods are often used within data science projects to discover previously unknown patterns in data.

2.3 Usage of AI in Production and Quality Processes

This paper focuses on the application of data science for quality processes in the scope of industrial production. This area includes all activities that influence the quality of the product and services provided to the customer of a manufacturing company. A wide bandwidth of factors contribute to this effort; ranging from individual quality checks on the product itself to the optimization of business processes that determine the execution of the technical and commercial processing.

Today, applications with practical relevance can be particularly found in the fields of predictive maintenance, process analysis and control, and forecasting of quality-related properties (Zhang et al, 2019).

3. APPLICATION EXAMPLES OF DATA SCIENCE TO IMPROVE QUALITY PROCESSES IN FOUNDRY SECTOR

3.1 Application Scenario: Foundry Sector

The foundry sector supplies essential basic components for many branches of industry. The production portfolio of Heidelberg Manufacturing Deutschland GmbH (HMD) ranges from printing machine components for its parent company to large components for wind power systems and smaller components for truck steering and braking systems. Compared to other manufacturing processes, the casting process is fundamentally dependent on the interaction of different divisions. Due to many divisions and processes involved, the variables influencing the cast components quality are very diverse.

To be able to produce a cast component, a sand mold is made by taking an impression of a pattern. This sand mold represents the positive image of the cast component to be produced. In order to reproduce the inner contours, so-called cores are inserted, which are also made of sand. To produce these cores, a core box is required, which represents the positive image of the core to be produced. In the next step, the molded sand form (which consists of an upper and lower part) is filled with liquid iron. The liquid iron is previously produced from raw materials in another division and kept warm at a predefined temperature until casting. After casting, the filled sand mold must be cooled down. Then the finished component can be extracted from the sand mold and further processed (shot blasting, heat treatment, etc.).

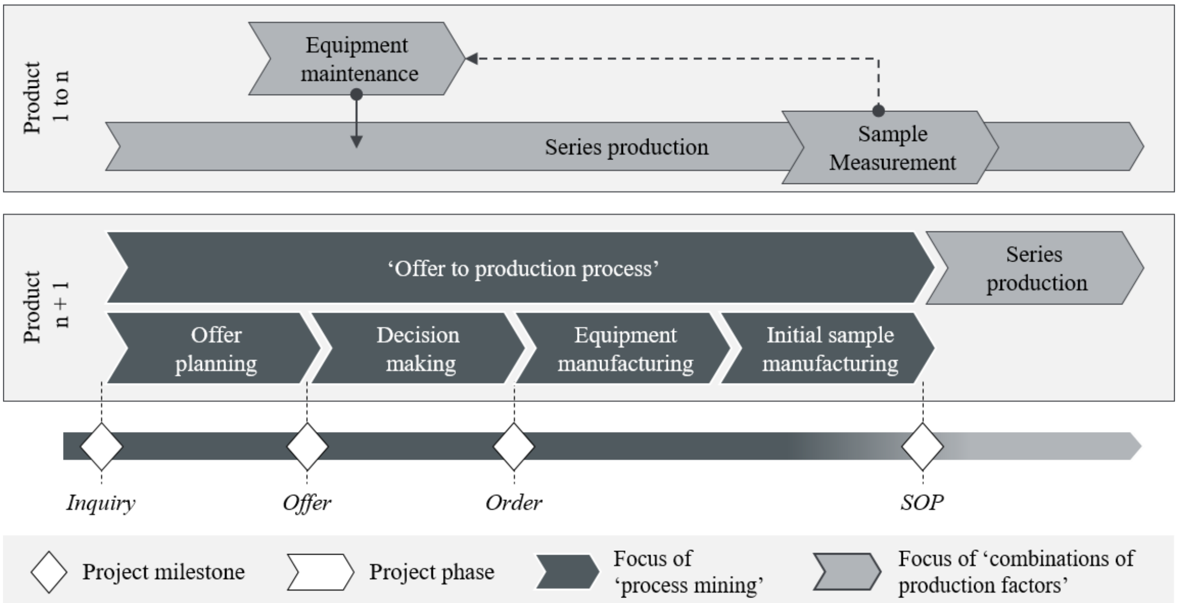

Figure 1. Considered quality processes

The interaction of all the processes is relevant for the quality of the cast component. In order to be able to produce new components, all process steps and manufacturing equipment (e.g. pattern, core box) must be planned. The so-called 'offer to production process' for new inquiries or change requests for cast components (cp. Figure 1) is accordingly dependent on the cooperation of various departments. In section 3.2 the optimization of this process with the help of the data science method process mining is discussed. At the end of a new inquiry with order, the component released by the series production release is produced in series (SOP). Series production usually refers to a multi-year cooperation in which agreed quantities are ordered on a regular basis. In general, minor changes are repeatedly requested. In some cases, products are also completely discontinued or replaced by substitute products. Each of these changes usually requires a new run through the ‘offer to production process’ in order to obtain the series release of the new or changed product.

Due to the constant contact of the manufacturing equipment with sand during series production, they wear out over time. Therefore, wear monitoring is required, in which the models and core boxes are regularly measured and, depending on the result of the measurement, revised so that their original condition is restored. Section 3.3 discusses which data science approach has been investigated for this wear monitoring.

3.2 Process Mining

The method process mining as one increasingly important subfield of data science has emerged strongly in recent years. The main concept of process mining is to gain information about real processes based on existing event data which is generated during the execution of processes (Berti et al, 2019). Whereas most methods in the field of data mining can be classified as data-centric, process mining is process-centric (Hand et al, 2001). During process mining, specific data mining algorithms are applied to event log data in order to identify trends, patterns or details contained in event logs recorded by an information system (Van der Aalst, 2016). One process mining method is the process discovery that aims to generate a process model exclusively based on log data.

<u>Relevant business process:</u>

The discovery technique is executed on existing event data from HMD. The goal of this implementation is to optimize the 'offer to production process’. The regarded process includes all process steps from a new inquiry for a product or product change request through planning, costing, offer negotiation and qualification up to the series release of the product. In the highly competitive international foundry industry, it is essential for a company to work in a customer-oriented manner. This also includes the fast and competent handling of new inquiries and changes. To improve in this field of customer interaction, the method process mining is applied.

<u>Implementation:</u>

The implementation of process mining requires preliminary work. As with most applications in the field of data science, data preparation and merging is essential for the application of process mining and represents the major part of the effort (Kiefer and Precht, 2019). The four major steps of the preparation process for the application of process mining are: selecting the right data (1), rating the individual data sources maturity level according to the Process Mining Manifesto (Van der Aalst, 2012) (2), generating a data model of the considered objects including their relations and the individual identification and attributes (3), merging the heterogeneous data sources (4).

The data, which has been produced through the preparation process, can then be played into the process mining tool. As a result, the process mining tool delivers the process model in a graphical manner and the according attributes e.g. the average duration of the process step.

The described application of process mining delivers a data based representation of the 'offer to production process’. This stated the basis for optimization measures to improve the customer satisfaction by optimizing the internal processes according to costs, time and quality. An extract of a process flow with the process mining method is shown in Figure 2.

<u>Conclusion:</u>

To sum up, the application of the data science method process mining has shown how processes can be optimized with the result of an improved customer satisfaction.

A prerequisite for the successful application of process mining is the quality of the data, which describe the process. In order to examine the data sources for suitability, a classification into maturity levels as stated in (Kiefer and Precht, 2019) is helpful. If a data source achieves a poor maturity level but is crucial for an analysis, the prerequisite for an analysis should be created in advance by revising/adapting the data source. Another prerequisite for the application of process mining is the underlying question. If the question is not aimed at an underlying target process, no useful result can be probably achieved.

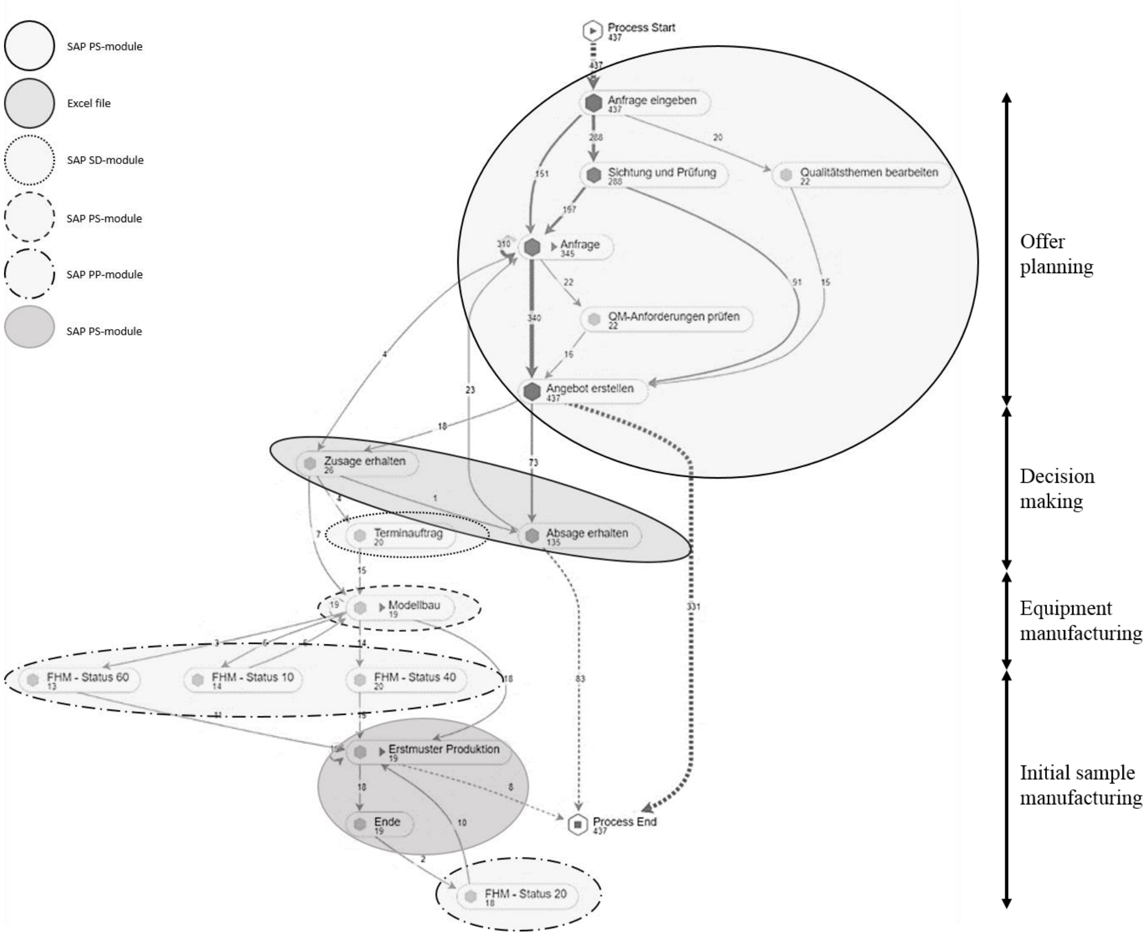

Figure 2. Schematic process flow generated by using process mining (Kiefer and Precht, 2019)

3.3 Combination of Production Factors

The goal of the second application of data science was the classification of different combinations of influencing parameters that lead to a finished casting. By classifying favorable and unfavorable pairings, favorable pairings should be systematically introduced to the production process and unfavorable pairings should systematically avoided.

Relevant business process:

In the foundry process flow, sand molds are produced with the help of patterns. Patterns are the positive of the finished component. In principle, a sand mold consists of a top box and a bottom box. At the same time, cores are made to reproduce internal contours or undercuts on the finished casting. The cores are inserted into the bottom box of the sand mold before the top box is placed above. Figure 3 shows the schematic structure of a molding box ready for casting. The molten iron is poured into the mold. Slow cooling of the liquid iron causes it to solidify. More information about the casting process can be find in (Beeley, 2001).

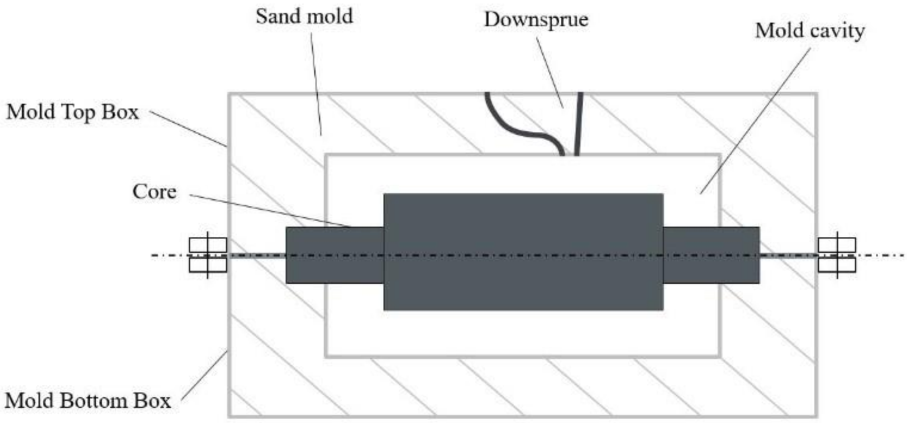

Figure 3. Schematic structure of a molding box

Influencing factors:

The two influencing parameters, core nest and model nest, are now considered in more detail. Depending on the dimensions and positioning, several parts can be cast simultaneously in one mold box. In order to be able to cast several parts per mold box, several individual models of the part are placed on a pattern plate (different pattern nests). At the same time, the cores are produced with the aid of a core box so that they can be inserted into the lower mold box later on. Several cores are produced simultaneously per one manufacturing process. The core box consists of multiple core nests. Each core nest and pattern nest is clearly marked. The finished casting contains information about the pattern nest it originated from (pattern nest number) as well as about the core it has been produced with.

Figure 4 visualizes the changes in sand mold and core over time due to wear on the model and core box. The dotted lines illustrate the development of the dimensions. These converge with increasing wear on the model and the core box. If the core has less and less play in the sand mold, it can move less and less in the mold during casting. This behavior should then become visible in the finished castings through a decreasing scatter of core-relevant dimensions.

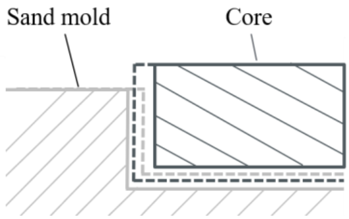

Figure 4. Wear effects on sand mold and core

The measurement data for determining this scatter comes from series production, where the finished castings are measured on a sample basis on a coordinate measuring machine. As portrayed in Figure 1, this measurement can result in the equipment (e.g. the pattern plate or the core box) having to be serviced. This happens, for example, when dimensional tolerance limits on the component are exceeded. The purpose of maintenance is to restore the equipment to its original or optimum condition.

Idea:

The dimensions of the sand mold and the core box converge. If there is no clearance between the sand mold and the core, stripping will occur. When the core is inserted, sand drifts off and lies at the bottom of the sand mold. During pouring, the loose sand is swept up and trapped in the finished casting.

By systematically identifying and classifying favorable (large clearance) and unfavorable (small clearance) pairings of pattern and core nests, with the help of data science methods, the service life of the pattern or core box can be extended by structurally inducing or preventing these pairings. This means that, in the optimum case, less maintenance of the equipment has to be carried out.

Implementation:

Measurement data of the finished castings serve as data basis. For each production batch, one casting per pattern nest is clamped onto the measuring machine. Defined points are then measured on this casting.

The scattering of core-relevant measurement data is then used to continuously classify the various combinations. The data analysis showed that it is not possible to make a statement or compare the different combinations on the basis of the scatter. Over a period of nine months, only a few measurements are available for the different combinations. An assessment of the current clearance between core and sand mold on the basis of the scatter of the measurements on the finished casting is not possible because the database is very limited. This also means that a comparison between the combinations is not possible.

Conclusion:

This use case cannot be used sensibly in the company. The database is very small compared to the number of different combinations. At the same time, data quality is limited by sometimes incomplete or ambiguous manual data entry, further limiting the usable database. No statement can be made about the quality of the different combinations, nor about whether there is a relevant difference between the combinations at all. What becomes clear here is that for the meaningful use of data science methods, the quality of the data basis is crucial.

This use case was initiated by a quality expert from the HMD in collaboration with an external IT expert with a scientific background. The IT side has to understand how the processes run in the company. The business experts from the company have to be able to objectively assess which IT methods and tools can do what, to what extent, and with what limits. For this, a deep mutual understanding is needed. At this point, it becomes clear that the application of data science in a company can only be used profitably if the competencies and knowledge from the various areas of data science (cp. Figure 1) are available and can be combined in a meaningful way.

4. RECOMMENDATIONS FOR ACTION

The experiences gathered by means of the presented use cases are used to formulate general recommendations for applying data science in manufacturing industry. Therefore, a separation into organizational aspects and technical ones is done in the following.

Today, many companies like HMD execute individual data science projects to gather experiences in this field. However, they often constitute isolated solutions that cannot prevail in the long term. To fully leverage the potentials of data science a fundamental data-driven culture needs to be anchored into the corporate strategy and all business processes. This requires the definition of business applications and their desired outcomes and the supply of required resources. Only then, comprehensive and long-term benefits can be derived that act self-reinforcing. Besides that, a successful implementation of the strategy is supported by interdisciplinary data science teams. They cover the required fields of expertise and collaborate with teams from the individual disciplines. This helps to overcome communication barriers between different subject matters and to disseminate basic data science knowledge within the company (Pricewaterhouse Coopers, 2020). The key task of this interdisciplinary team is to verify that a suitable data science method is picked regarding the desired outcome.

On the technical dimension the provision of data with the required degrees of quality and quantity is key. Both presented use cases at HMD have revealed major deficiencies in this regard. The current situation in industry is often characterized by insufficient amounts of data that would be required to extract the desired knowledge. Beyond that, existing datasets are usually spread over incoherent data sources that don’t apply standardized data formats. Thus, an automatic processing isn’t possible by default. The formatting, sorting, and merging of data as pre-task can require a disproportionate effort. Thus, a proper foundation has to be established first by providing the required amounts and maturity of the data and connecting the individual data sources (Kiefer and Precht, 2019). Furthermore, the datasets should include reference parameters so that its origin and semantic content is defined. For example, a raw dataset of temperature measurements is meaningless unless it is defined in which location the temperature has been measured and which measurement belongs to which product or batch. Each data science project should start with a review of these aspects that determines if the requirements are fulfilled or if substantial work has to be put in first.

5. SUMMARY AND OUTLOOK

This paper illustrates how data science methods can be usefully and profitably used in manufacturing industry taking the example of the foundry sector. Focus of these applications are production-related engineering and manufacturing processes with regard to the topics quality and process management.

The main research question of this paper is the investigation of potentials, limits and challenges of using data science methods to improve quality processes in operational practice. In terms of a socio-technical overall view, not only methodical issues but also information-technical and organizational aspects are discussed. The gained knowledge particularly results from studies of two processes: the ‘offer to production process’ and ‘combination of production factors’. The collected findings ultimately lead to recommendations for action, which are intended to give companies a kind of framework for a meaningful economic use of data science methods in operational practice.

Although much progress has already been made in the field of data science and quality processes so far, the following topics remain to be addressed in further research activities:

- Development of a standards-based, semantic quality data and quality service model that transparently provides all relevant quality-related development, planning, production and customer data (and functions) to the identified applications within the quality process chain
- Development and use of effective and efficient AI algorithms for continuous planning and control of quality processes along the entire product life cycle

REFERENCES

Beeley, P., 2001. Foundry Technology. *Oxford*, Butterworth-Heinemann.

Berti, A. et al, 2019. Process Mining for Python (PM4Py): Bridging the Gap Between Process and Data Science. *In arXiv:1905.06169*.

EIT Digital, 2020. Overview on maturity of AI innovations in manufacturing. *Study*.

Grus, J., 2019. Data Science from Scratch. *O’Reilly Media, Inc.*, Second Edition.

Hand, D.J. et al, 2001. Principles of data mining. *In Cambridge, Mass: MIT Press*.

IFS, 2018. Künstliche Intelligenz und maschinelles Lernen. *Study*.

Kiefer, J. and Precht, M., 2019. Usage of process mining in the ‘offer to production process’ of a contract manufacturer for cast components. *Proceedings of the 4th International Conference on Big Data Analytics, Data Mining and Computational Intelligence – BIGDACI 2019*.

Kotu, V. and Bala, D., 2018. Data Science – Concepts and Practice. *Morgan Kaufmann*, Seond Edition.

Ozdemir, S. and Sunil, K., 2018. Principles of Data Science. *Packt Publishing*, Second Edition.

PAiCE Scientific Assistance, 2018. Potential of artificial intelligence in Germany’s producing sector. *Study on behalf of the Federal Ministry for Economic Affairs and Energy (BMWi)*.

Pricewaterhouse Coopers, 2020. An Introduction to Implementing AI in Manufacturing. *White Paper*.

Shi, H. et al, 2020. Construction of Industrial Internet of Things Based on MQTT and OPC UA Protocols. *In 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*.

Van der Aalst, W., 2012. Process Mining Manifesto. *In Business Process Management Workshops*, Springer Berlin Heidelberg.

Van der Aalst, W., 2016. Process Mining: Data Science in Action. *In 2nd ed. Imprint: Springer*, doi: 10.1007/978-3-662-49851-4.

Wahlster, W., 2021. Ten Years of INDUSTRIE 4.0. *DFKI Interview with Prof. Dr. Wolfgang Wahlster, published in DFKI NEWS*, Issue 1/2021.

Zhang, W. et al, 2019. Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. *IEEE Systems Journal 13*, pp. 2213-2227.