Industrial Data Science - Interdisciplinary Competence for Machine Learning in Industrial Production

Lukas Schulte¹ [0000-0002-0613-7927], Jacqueline Schmitt¹ [0000-0002-5780-702X],
Lukas Stankiewicz¹ [0000-0003-0155-5932] and Jochen Deuse^{1,2} [0000-0003-4066-4357]

¹ Institute of Production Systems, TU Dortmund University, Leonhard-Euler-Straße 5,
44227 Dortmund, Germany

² Advanced Manufacturing, School of Mechanical and Mechatronic Engineering,
University of Technology Sydney, Australia

lukas.schulte@ips.tu-dortmund.de

Abstract. Within the increasing digitalization, the widespread application of modern information and communication technologies and the technological ability to systematically and comprehensively capture and store data allow to build data storages of unprecedented size and quality. The evaluation and efficient use of the implicit knowledge in the data to support decision-making processes is becoming increasingly significant in manufacturing companies. Thus, new requirements arise for the qualification and competence development to efficiently solve engineering applications and issues in manufacturing and assembly with advanced data-driven methods. This paper presents the contribution of a qualification concept for Machine Learning in industrial production that has been realised within a recent research project funded by the Federal Ministry of Education and Research. This concept has been designed and validated within the university curriculum for graduate students in mechanical and industrial engineering, computer science and statistics. Taking into account the current challenges in manufacturing and assembly, the contribution of this enhanced interdisciplinary competence development can be considered quite significant. The results, findings, and future enhancements are presented within this paper.

Keywords: Competence Development, Industrial Data Science, Industrial Manufacturing, Interdisciplinary Competence, Machine Learning

1 Introduction

The German manufacturing sector employs eight million people, making Germany one of the major business and industrial locations worldwide. However, the production environment is subject to constant change through ever evolving trends and challenges. Alongside high productivity and profitability, a highly diverse and customizable product portfolio has become a central strategic competitive factor for manufacturing companies [1]. This, however, entails more flexible and increasingly complex manufacturing and assembly systems, pushing traditional methods and techniques progressively to their limits.

At the same time, the increasing digitalization and the technological ability to systematically and comprehensively capture and store data allow to build data storages of unprecedented size and quality [2]. This data, however, can be used for optimization and decision-making processes within production. By discovering non-trivial and yet unknown structures and correlations, Machine Learning (ML), as an interdisciplinary subject of computer science and statistics, provides intelligent and automated processing of large amounts of data [3]. While data analysis alone, however, contributes to solving practical problems only to a limited extent, it represents, in particular in connection with the domain-specific expert knowledge, a future success factor of the companies [4]. Therefore, methodological competence in the field of ML as well as a good technical understanding of practical engineering questions are required [5]. This intersection of computer science, statistics and engineering constitutes the field of Industrial Data Science (IDS).

However, companies often lack the resources to adequately assort such interdisciplinary teams for IDS projects. In order to solve this dilemma, the qualification of academic graduates and continuing education of industrial professionals must be modified accordingly. [5] Therefore, the research project “InDaS”, funded by the Federal Ministry of Education and Research in the program “Qualification programs and research initiatives in the field of Machine Learning”, focuses on the development and validation of a comprehensive and interdisciplinary qualification concept of methodical competences in the field of ML with practical applications in industrial production.

2 State of the Art

2.1 Machine Learning in Production

Machine Learning (ML) has proven to be very advantageous in various fields of application. Thereby, the success is encouraged by the invention of increasingly sophisticated ML models [6], the availability of large data sets [7] and the development of software platforms [8] which enable an easy employment of vast computational resources for the training of ML models on large data sets.

In industrial production, ML methods can be used to overcome existing problems and boundaries in a number of application scenarios along the production process chain. Successful applications can be found for various tasks in production planning and optimization [9], quality improvement and prediction [10, 11], predictive maintenance [12] and energy efficiency optimization [13]. The data used ranges from master data, over transaction, log and sensor data to text, voice, video and audio data [14]. Algorithms applied include supervised and unsupervised methods such as tree-based classifiers, Support Vector Machines (SVM), Neural Networks, Generalized Linear Models (GLM), Clustering techniques, etc. [15].

Nevertheless, the existing research gap is a lack of interdisciplinary competence that allows the appropriate formulation of learning tasks and application of algorithms for engineering questions in a situations-oriented manner.

2.2 Educational Concepts

As mentioned above, an increasing technical and functional competence to handle large amounts of data in production in the era of Industry 4.0 is needed [16]. In this context and in conjunction with the fact that learning is not directly observable, a multitude of divergent theories, models, and approaches emerged to explain human learning [1]. Educational concepts that focus on knowledge generation and transfer for individual target groups are equally diverse.

Practically-oriented approaches such as learning factories have become highly promising for knowledge and innovation transfer in academic as well as professional education [17]. Teaching and learning is based on actual problem situations and issues from research and industry as well as on practice-oriented case studies and defined didactic methods [18]. Thereby, the interconnection of research, industry and education enables the establishment of interdisciplinary, multi-dimensional teaching and learning experiences. As a result, a sustainable transfer of knowledge, the development of competence and the ability of creative and self-organized action can be captured [18, 19].

This kind of approach is also required in order to transfer the essential knowledge and skills of IDS to industrial manufacturing and assembly. Through a strong practical orientation, the ability to apply adequate methods to real questions is trained in addition to the mere imparting of knowledge. There is a comprehensive range of seminars, trainings and certificates for various fields of ML including visual analytics, data preparation and data quality, ML algorithms and methods as well as others [20–25]. Nevertheless, each of these qualification measures is based on a specific methodological topic and does not enable the attendees to experience the extent of actual industrial ML projects. **Table 1** shows a comparison of different lectures for Data Science and their focuses.

Table 1. Comparison of Data Science seminars concerning content alignment

Lecture	Institution	Data Management	Data Preparation and Visualization	Machine Learning Tools	Machine Learning Methods	Industrial Application	Engineering Domain Knowledge	Practical Sessions	Real Use Case Processing
Data Mining and Machine Learning	TU Darmstadt	○	◐	◐	●	○	○	◐	○
Data Mining	Universität Hamburg	○	●	◐	◐	○	○	◐	○
Data Mining	University of Hildesheim	○	●	◐	◐	○	○	◐	○
Machine Learning	Leipzig University	○	○	○	●	○	○	○	○
Machine Learning	TU Dortmund University	○	○	◐	○	○	○	◐	○
Data Science	htw Saar	○	◐	◐	●	○	○	○	○
Knowledge Discovery in Databases	TU Dortmund University	○	○	◐	○	○	○	◐	○
Big Data Analytics	University of Hildesheim	●	○	◐	◐	○	○	◐	○
Digitization and Data Analytics	TU Dresden	◐	○	◐	●	○	○	◐	○
Big Data Analytics	Fraunhofer Big Data AI	◐	○	◐	◐	○	○	◐	○
Machine Learning Master	RapidMiner	○	●	◐	◐	○	○	◐	○
Data Mining	Prokoda	○	●	◐	◐	○	○	◐	○
Big Data Analytics and Data Mining	IHK Akademie Schwaben	○	○	○	◐	◐	○	◐	○
Data Science	Harvard University (edx)	○	●	◐	◐	○	○	◐	○
Data Science	IBM (edx)	●	●	◐	○	○	○	◐	○
Industrial Data Science	TU Dortmund University	●	●	◐	●	●	●	●	●

consideration
○ none
◐ rudimentarily
◑ partly
● mostly
● comprehensive

Table 1 shows that, with regard to the seminars offered on the subject of Data Science, there is little or no reference to industrial application and the introduction of engineering domain knowledge. In addition, the consideration of data management and the processing of real use cases is only occasionally part of the seminars.

3 Concept and Development

In order to sustainably strengthen the application and dissemination on ML in industrial production, competence development and enhancement may not be limited to industrial professionals, but address young graduates in university education in particular. Thereby, qualification requirements arise from current trends and developments in industry as well as the state of knowledge and skills of professional experts. To design the education and competence enhancement as efficiently and practice-oriented as possible, a mature concept taking all given requirements fully into account is a necessary precondition.

3.1 Qualification Requirements for Industrial Data Science

Future concepts for qualification and competence development are exposed to increased requirements and must be adapted and supplemented by additional aspects in order to guarantee sufficient competencies in the age of digitalization. The rising complexity of industrial processes requires an increase in specialist competencies, while the greatest challenge is to transfer theoretical knowledge to practical questions and areas of application [5]. Therefore, new approaches should include practical applications and experimentation so that a comprehensive understanding can be created.

A survey by Bauer et al. in 2018 revealed, that only 5 out of 57 companies (9%) within different industrial sectors such as automotive, electronics and mechanical engineering, already use ML widely [5]. Further 27 companies (47%) stated that they had gained initial experience or used methods on a small scale. The remaining 25 companies (44%) have not yet used ML, but plan to do so in the future.

Thereby, the lack of corresponding methodological competence has been identified as the major obstacle for the use of ML. To overcome this shortcoming, a catalogue of requirements must be fulfilled by adequate qualification and competence development concepts for Industrial Data Science [5]. Based on didactic-methodological knowledge transfer, the subjects of data management, ML and domain knowledge are the threefold foundation of IDS education. These core topics can be illustrated by three overlapping thematic areas (see **Fig. 1**).

Fig. 1. Concept for qualification and competence development of Industrial Data Science as part of the research project "InDaS" [5]

3.2 Conception of interdisciplinary competence development

Within the research project "InDaS", the aforementioned requirements are addressed in a strongly practice-oriented and interdisciplinary qualification concept for academic graduates and industrial professionals. This concept combines engineering knowledge and skills with methodological expertise in statistics and computer science. Through close cooperation with manufacturing companies and an early integration and consideration of practical requirements and applications, a high practical relevance of taught contents is sought. The processing of industrial use cases by students with supervision by industrial professionals and university lecturers represents a new approach to teaching. Students work on problems in a real context and realistic levels of difficulty and make contact with companies during their studies. The educational offer addresses students aiming for a master's degree in mechanical and industrial engineering, computer science or statistics as well as industrial specialists. The course is divided into a theoretical phase to first convey basic contents and a practical phase to apply gained knowledge to practical use cases and questions.

3.3 Development of Methodological Competence

The basic theoretical contents have to be imparted to enable an independent processing of industrial questions with appropriate methodologies. Hereby, the selected theoretical contents are determined by the specific needs of industry to react to current trends and changes. With regard to the processing of large data sets, many companies criticize the manual acquisition and processing of unstructured data and the data administration [26]. Thereby, especially large data volumes are challenging and a faster data supply is desired. Further identifiable problems include a lack of data cleansing and analysis capabilities, a non-agile business intelligence infrastructure, a lack of access to data sources, a poorly intuitive user interface and errors in data management. [27]

To address the before mentioned issues and requirements, the curriculum comprises the thematic modules of data management, machine learning and engineering domain knowledge, built on the basis of didactic-methodical knowledge transfer (see **Fig. 2**).

Fig. 2. Contentual and didactical design of the IDS qualification concept

In the theoretical part, complex contents are reduced to their general idea in order to create a basic understanding of multiple methods and approaches. In addition, potential fields of application, problems and limits of learned methods are discussed to enable the students to take them into account in the later practical implementation. Within the experience-based learning approach, learning takes place in interdisciplinary working groups with the inclusion of industrial application cases as well as a project-oriented development of individual solutions.

The efficient management of data requires knowledge in the areas of database systems, data administration, resource management and cleansing and processing of structured and unstructured data. To enable the students to practically realize these tasks, experiences with cloud solutions, database concepts, and script languages, e. g. Python, R or RapidMiner, for modelling and programming data science solutions are taught.

In addition to mere knowledge of different ML methods and approaches, their demand-oriented application is an essential success factor and is therefore explicitly taken into account in IDS qualification. Contentually, the processing pipeline of data preprocessing, modelling, validation and evaluation is covered through the focal points of data visualization, feature engineering, ML algorithms, validation and evaluation.

From the engineering point of view, it is essential to understand the manufacturing and assembly processes to detect the potential of ML for individual business units. To execute ML projects systematically, respective process models such as CRISP-DM (cross-industry standard process model for data mining) [28] should be known. Additionally, domain-specific structuring models, e. g. the product life cycle, are needed to examine the achieved results for their relevance and cost-effectiveness regarding the underlying business case.

Based on the mentioned modules, the students are capable of solving industrial tasks with data-driven methods and developing innovative solution and improvement approaches in manufacturing and assembly. They are able to lead and support interdisciplinary project teams and to pass on their interdisciplinary competences and skills.

3.4 Transfer to Practical Applications

It is an essential success factor to ensure that young academic graduates and industrial professionals not only know the theoretical backgrounds of IDS, but also build up a deep understanding and are able to apply methods situatively and correctly. Therefore, the second part of the qualification concept consists of the practical processing of industrial case studies in interdisciplinary working groups. Under the involvement of industrial specialists, different case studies covering ML applications such as pattern

recognition and image classification (see **Fig. 3**) enable the enrichment of theoretical knowledge with practical experience through knowledge transfer and experience-based learning.

Provided data formats	Objective	Methods applied
Automatic quality evaluation of injection molded parts	Classification of defective injection molded parts
Classification and control of chemical batch processes	Evaluation, forecasting and optimization of process behaving
Identification of field failures based on EOL measurements	Identification of potential returns and responsible parameters
Reduction of pseudo faults in the control of plug connection	Automatic classification of incorrectly assembled connectors

Fig. 3. Case studies of the 2019 practical IDS seminar at TU Dortmund University

After the general introduction of the use cases with support of industrial process experts, the project takes place in form of a three-month self-organized, independent group work with regular scientific supervision by industrial and methodological experts. The groups compose of graduate students from the disciplines of statistics, computer science as well as mechanical and industrial engineering, ensuring interdisciplinary key competencies. The results are presented in a joint seminar, including a discussion of results and outlook to further research activities.

4 Concept Evaluation

The evaluation of the concept is based on the mandatory course evaluation as well as the results of the learning progress assessment.

A 60-minute written examination at the end of the theoretical phase serves as a learning success control. The imparted topics of data administration, statistical basics of ML, ML methods and engineering domain knowledge are queried in one assignment each, consisting of several subtasks. The degree of complexity ranges from simple inquiry of knowledge, the transfer onto new facts and conditions and the linkage of information to solve more complex problems. The results of the examination and the results of the anonymous evaluation in the form of a written questionnaire point to identical findings.

The results and feedback from students in all three subject areas show that the comprehensibility of the contents from the individual disciplines has been of a similar high standard. As a result, it can be concluded that the preparation of the content can be regarded as quite successful, taking into account the heterogeneous target group [29]. The only criticism received has been directed to the introduction to the programming language R, which has been perceived as too condensed by students without prior knowledge. Great interest in the IDS education has been reflected among the university students as well as on the side of industrial partners. As a key figure, the number of participants which increased from 51 in the winter term 2018/19 to 71 in the winter

term 2019/20 (+39%) can be mentioned. A survey among the cooperating industry partners has revealed that all of them rate the achieved results as above their expectations and are willing to cooperate again in the next practical seminar.

5 Conclusion and Outlook

Enabled by the increasing digitalization, the widespread application of modern information and communication technologies and the availability of large amounts of data, ML applications are becoming increasingly significant in industrial manufacturing and assembly. To address the associated qualification requirements, new educational concepts have to be developed. The “Industrial Data Science” qualification concept enables an interdisciplinary competence enhancement on the intersection of computer science, statistics and the engineering domain. Through the combination of teaching of theoretical contents and the transfer to practical applications, a comprehensive and practically-oriented qualification is facilitated. The learning success and feedback from the concept evaluation in university education at TU Dortmund University has shown that the course addresses current industrial questions and requirements and is subject to high interest. With the goal of integrating potential benefits into manufacturing companies sustainably, the qualification concept has to be expanded and continuously improved in order to guarantee ongoing actuality. Therefore, the qualification in IDS must be anchored in the university curriculum for graduate students of relevant disciplines such as statistics, computer science and especially engineering and the target group has to be expanded onto skilled professionals in manufacturing companies.

Acknowledgment Part of this work has been funded through the research project “InDaS” (BMBF, funding code 01|S17063A) by the Federal Ministry of Education and Research in the context of “Qualification programs and research initiatives in the field of Machine Learning” program.

References

1. K. Weisner, “Contribution of the development of individual competence in dealing with a wide range of variants by adapting motoric learning theories”, 1st ed. Herzogenrath: Shaker, 2019.
2. M. Eickelmann, M. Wiegand, B. Konrad and J. Deuse, “The importance of Data Mining in the context of Industry 4.0” (in German) in *Zeitschrift für wirtschaftlichen Fabrikbetrieb (ZWF)*, vol. 110, no. 11, pp. 738–743, 2015.
3. U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, “From Data Mining to Knowledge Discovery in Databases” in *AI Magazine*, vol. 17, no. 3, pp. 37–54, 1996.
4. J. Deuse, O. Erohin and D. Lieber, “Knowledge discovery in networked, industrial databases” (in German) in *Industrie 4.0: Wie intelligente Vernetzung und kognitive Systeme unsere Arbeit verändern*, Hamburg, 2014, pp. 373–395.

5. N. Bauer, L. Stankiewicz, M. Jastrow, D. Horn, J. Teubner, K. Kersting, J. Deuse and C. Weihs, "Industrial Data Science: Developing a Qualification Concept for Machine Learning in Industrial Production" in *European Conference on Data Analysis (ECDA)*, Paderborn, 2018.
6. V. Mnih, N. Heess, A. Graves and K. Kavukcuoglu, "Recurrent models of visual attention" in *Advances in neural information processing systems*, 2014, pp. 2204–2212.
7. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla and M. Bernstein, "Imagenet large scale visual recognition challenge" in *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.
8. T. Chilimbi, Y. Suzue, J. Apacible and K. Kalyanaraman, "Project adam: Building an efficient and scalable deep learning training system" in *11th Symposium on Operating Systems Design and Implementation*, 2014, pp. 571–582.
9. R. Wallis, O. Erohin, R. Klinkenberg, J. Deuse and F. Stromberger, "Data mining-supported generation of assembly process plans" in *Procedia CIRP*, vol. 23, pp. 178–183, 2014.
10. J. Deuse, J. Schmitt, J. Bönig and G. Beitingner, "Dynamic X-Ray inspection in electronics production: Use of Data Mining methods for quality inspection" (in German) in *Zeitschrift für wirtschaftlichen Fabrikbetrieb (ZWF)*, vol. 114, no. 5, 2019.
11. J. Deuse, J. Schmitt, M. Stolpe, M. Wiegand and K. Morik, "Quality prediction for bottleneck relief in injector production using Data Mining" (in German) in *Gronau (Hg.) 2017 – Industrial Internet of Things*, 2017, pp. 47–60.
12. P. Strauß, R. Wöstmann, M. Schmitz and J. Deuse, "Enabling of Predictive Maintenance in the Brownfield through Low-Cost Sensors, an IIoT-Architecture and Machine Learning" in *2018 IEEE International Conference on Big Data*, Seattle, 2018.
13. M. Peña, F. Biscarri, J. I. Guerrero, I. Monedero and C. León, "Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach" in *Expert Systems with Applications*, vol. 56, pp. 242–255, 2016, doi: 10.1016/j.eswa.2016.03.002.
14. M. Weber, S. Dehmelt, M. Shahd and K. Hampe, "Potentials and use of Big Data: Results of a representative survey of companies in Germany" (in German) in *Hg. v. Bitkom*, 2014.
15. A. K. Choudhary, J. A. Harding and M. K. Tiwari, "Data mining in manufacturing: A review based on the kind of knowledge" in *Journal of Intelligent Manufacturing*, vol. 20, no. 5, pp. 501–521, 2009, doi: 10.1007/s10845-008-0145-x.
16. M. Achenbach, L. Schulte, J. Deuse and P. Buhr, "Successful and instructive at the same time: Competence development for dealing with system complexity through joint learning by student and industry employees" (in German) in *Industrie 4.0 Management*, vol. 32, no. 3, pp. 11–14, 2016.
17. E. Abele, G. Chryssolouris, W. Sihn, J. Metternich, H. ElMaraghy, G. Seliger, G. Sivard, W. ElMaraghy, V. Hummel, M. Tisch and S. Seifermann, "Learning

- factories for future oriented research and education in manufacturing” in *CIRP Annals*, vol. 66, no. 2, pp. 803–826, 2017, doi: 10.1016/j.cirp.2017.05.005.
18. J. Enke, J. Kaiser and J. Metternich, “The learning factory as export success” (in German) in *Zeitschrift für Wirtschaftlichen Fabrikbetrieb (ZWF)*, Carl Hanser Verlag, München, vol. 112, no. 10, pp. 644–647, 2017, doi: 10.3139/104.111806.
 19. M. Tisch, C. Hertle, J. Metternich and E. Abele, “Measuring learning success in learning factories - competence-oriented further development of practical training” (in German) in *Industrie Management*, GITO Verlag, Berlin, vol. 30, no. 3, pp. 20–24, 2014.
 20. Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme (IAIS), *Methodenspezifische Schulungen: Schulungsangebote*, <https://www.bigdata.fraunhofer.de/de/datascientist/seminare.html##>, last accessed 2020/01/16.
 21. RapidMiner, *RapidMiner Training*, <https://rapidminer.com/learn/training/##>, last accessed 2020/01/16.
 22. PROKODA GmbH, *Seminare*, https://www.prokoda.de/seminare/seminarliste/index_ger.html?Seminar=Data%20Mining&ortstringalle=Alle&dFrom=1&mFrom=1&yFrom=2015&dTo=31&mTo=12&yTo=2015&mitDatum=1&Garantie=%25&Seite=1&submit=Seminartermine+suchen##, last accessed 2020/01/16.
 23. IHK Akademie Schwaben Weiterbildung GmbH, *Big Data Analytics und Data Mining*, <https://www.ihk-akademie-schwaben.de/weiterbildung/seminar/226768/augsburg/big-data-analytics-und-data-mining##>, last accessed 2020/01/16.
 24. IBM, *Data Science Professional Certificate in Professional Certificate in IBM Data Science*, <https://www.edx.org/professional-certificate/ibm-data-science#>, last accessed 2020/01/16.
 25. Harvard University, *Professional Certificate in Data Science*, <https://www.edx.org/professional-certificate/harvardx-data-science##>, last accessed 2020/01/16.
 26. R. Göbel and J. Weinfurter, “Guide: Industry 4.0 and Smart Data. The world of large volumes of data in companies” (in German) in *Hg. v. eBusiness-Lotse Oberfranken*, 2015.
 27. A. Lenk, “Smart-Data-Business – 10 thesis on the use of Big Data solutions in business: Short study of Smart-Data accompanying research in cooperation with the Business Application Research Center (BARC) and the Bundesverband der IT-Anwender e. V. (VOICE)” (in German) in *Smart-Data-Begleitforschung*, 2015.
 28. P. Chapman, J. Clinton, R. Kerber, T. Khabaz, T. Reinartz, C. Shearer and R. Wirth, “CRISP-DM 1.0: Step-by-step data mining guide” 2000.
 29. J. Deuse, K. Weisner, J. Schmitt and J. Schallow, “Innovative models and methods of industrial engineering for designing human work” (in German) in *ERKENNEN.LERNEN.VERÄNDERN.: Die Arbeit des Menschen in der digital vernetzten Welt*, Stuttgart, 2019.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.