A learning factory approach on machine learning in production companies

How a learning factory approach can help to increase the understanding of the application of machine learning on production planning and control tasks.

Alexander Rokoss, Kathrin Kramer, Matthias Schmidt

1. Introduction

Technological progress and increasing digitalization offer many opportunities to production companies, but also continuously confront them with new challenges. On the one hand, automation of processes is progressing in manufacturing and upcoming technical support systems, such as automated guided vehicles, are leading to significant changes in workflows. On the other hand, large amounts of work within companies are still done by humans. This is also the case with production data processing. Although in many enterprises data is already collected and sorted automatically, the final evaluation of this data and especially decision-making is often done manually. Especially decisions that cannot be handled on the basis of conditional programming are usually made by humans.

The use of artificial intelligence (AI), in particular the use of machine learning (ML) algorithms, represents a promising approach to support the automation of complex decisions (Buxmann/Schmidt 2019; Kuhn/Johnson 2013). A steep increase of scientific publications in recent years (Schmidt et al. 2020) evidences the trend that more and more companies and institutions are dealing with the utilization of ML in production planning and control (PPC). However, many of the publications originate from a purely scientific environment, so a widespread use in companies is not visible yet (Usuga Cadavid et al. 2020). Rammer et al. state in their report that with 17,500 analyzed companies, 5.8% are already using AI in products, services, or internal processes. At the same time, the report indicates that 43% of the overall 22,500 AI-related vacant jobs at the time of the survey, could not be staffed (Rammer et al. 2020). The resulting massive shortage of skilled workers in the field of AI must be addressed in the short- and mid-term by training and educating existing employees in production companies. A promising approach to build up competencies within production companies is the use of learning factories as a knowledge transfer enabler. Learning factories offer participants the opportunity

DOI: https://doi.org/10.30844/wgab_2021_8

to apply new methods in a realistic environment without the risk of negative consequences on live production processes (Abele et al. 2017). In this context, the influences of applied methods can be directly experienced without time delay, resulting in better learning results compared to conventional full-frontal teaching methods (Sackey et al. 2017). Supplementary, a gamification of learning content eases the learning process thanks to facilitating a "learning by doing" mentality (Keepers et al. 2020).

As mentioned above, professionals in PPC daily face the task of decision making on complex issues. The decisions are often based on experiences, sometimes supplemented by indicators generated from existing data. As early as 1992, Yang et al. proposed the use of intelligent systems for planning production processes (Yang et al. 1992). To be able to make decisions more reliable and to preserve the knowledge of long-time employees, intelligent support systems using ML methods are needed. To develop and operate such systems, professionals need specific training that enable them to acquire the necessary skills and competencies. Literature shows that the focus of most learning factories, operated at corporate or university level, is on lean management and demonstrating digitalization technologies to support lean manufacturing (Martinez et al. 2020). Although ML approaches have already been implemented, PPC tasks are very rarely addressed (Martinez et al. 2020). Grounded on their relevance for meeting contracted delivery dates with the customer, one central component of the planning process in PPC is the determination of lead times (LT). A possible procedure for the application of ML methods is the use of a data mining process like the Cross Industry Standard Process for Data Mining (CRISP-DM) (Chapman et al. 2000; Usuga Cadavid et al. 2020). The combination of the general approach of CRISP-DM combined with the specific field of ML-supported LT prediction in a learning factory environment has not yet been deeply investigated. A learning factory training on the use of CRISP-DM and the utilization of ML to determine production LT in a job shop can help to improve the knowledge transfer on developing ML applications in production environments.

Henceforward, this book chapter describes a learning factory training on the utilization of ML approaches in PPC. A brief introduction into learning factories and ML in PPC is followed by the design methodology for the learning factory training. Afterwards necessary competencies are identified, leading into the conception of the training phases. Lastly, the conclusions are summarized, and future research possibilities are outlined.

2. Learning factories

Learning factories allow participants to actively engage in the implementation and improvement of production processes in a realistic environment. One of the key advantages is the direct response of the production system as a result of decisions made by the participants without time delay (Sackey et al. 2017). Depending on the didactic purpose, learning factories teach a specific subfield of production. To equip learners with tools for immediate reaction, the tasks to be solved always derive from problems occurring in real world production systems (Abele et al. 2019). This results in participants developing professional competencies while avoiding to harm the real world production environment (Abel et al. 2013). Another major advantage of learning factories is the superior performance in generating knowledge compared to traditional teaching (Cachay et al. 2012). Physically experiencing a learning process promises a significantly deeper anchoring of the new information as well as more joy in the learning process itself. The lack of consequences of errors for crucial processes in the live production environment clearly encourages curiosity and the motivation to experiment in learning factories (Deslauriers et al. 2011; Haghighi et al. 2014).

The majority of existing learning factories currently address topics of lean management as well as how to deal with digital transformation technologies (Abele et al. 2015; Sackey et al. 2017; Veza et al. 2017). In a recently published study Martinez et al. show the rapid increase in publications dealing with learning factories since 2015 (Martinez et al. 2020). Lately ML methods have already been implemented in learning factories addressing predictive maintenance (Daniyan et al. 2020), quality control (Oberc et al. 2020), activity and position recognition (Hofmann et al. 2020; Zhang et al. 2020) or process mining on end-to-end order processing (Schuh et al. 2020). For the training presented here products and the environment of the learning factory already existing at the Leuphana University of Lüneburg are used, allowing access to the ecosystem data and IT systems.

3. Machine learning in production planning and control

As most tasks in PPC are depending on predictions in complex environments, the selection between different actions turns out to be difficult (Nyhuis/Wiendahl 2006; Wiendahl 2014). ML approaches could help to solve those problems. Arthur Samuel (1959) defines ML as the field of study in which not every step is programmed to the computer by humans. Instead computers learn independently by developing rules (Chollet 2018). Using labeled training datasets, the rules mentioned are determined and applied to an unknown dataset.
Since 2016 a steady growth in publications related to ML can be observed (Döbel et al. 2018). A bibliometric analysis by Schmidt et al. (2020) shows that the distribution of ML's potential across main PPC tasks (Schäfers/Schmidt 2015) in scientific publications is extremely heterogeneous and varies from no publication to more than one third of all publications classified. One of the main focus areas of publications classified by Schmidt et al. were LT scheduling and scheduling in general, whereas in previous years the focus was on sequencing. This trend could be explained either by the large share of already existing approaches (e.g. case studies

by Gyulai et al. 2018; Lingitz et al. 2018) or by the development progress in the ML area, resulting in easier applicability to problems that are more complex. As publications in recent years have increasingly focused on scheduling and LT determination, the following section takes a deeper look into the principles of LT determination using regression models.

4. Methodology for the design of the learning factory training

The desired outcome at the end of the learning factory training is the knowledge on how ML methods can be used for the determination of LT for new production orders in a production environment. As a data mining process like CRISP-DM offers a structured procedure for handling the required data and making an accompanying documentation, it can function as the basis for the learning factory training. According to CRISP-DM, it records relevant production data, gains transparency of the processes from the data obtained and finally generates recommendations for action utilizing ML (Chapman et al. 2000). Within the learning factory training, participants will encounter the difficulty of precisely determining LT under uncertainty. On the one hand, the training guides the learners through the data collection process using available digitalization technologies. On the other hand, it encourages participants to setup a ML application themselves to determine LT. By going through CRISP-DM stepwise, participants are enabled to setup a ML application on LT determination that is trained and tested with a dataset originating from a real job shop production site. The methodology shown in Figure 1 is used to develop the required training phases.

The conceptual design of the learning factory training begins with the definition of teaching objectives. Next, associated competencies need to be identified. Three steps are involved in identifying the competencies. First, the underlying theory on CRISP-DM as well as an example of ML methods used for LT determination in practice is analyzed regarding the definition of teaching objectives. Second, necessary competencies on setting up CRISP-DM using ML methods are extracted from the given theory. In a third step, findings are then compared with Danyluk/Bucks analysis of AI competencies on data science and Long/Magerkos 2020 paper on AI literacy. The analyzed competencies are recorded using the matrix structure for evaluating technical-methodical competencies according to Abel et al. (2013). Next, a fictitious scenario is developed, that forces learners to deal with the determination of LT within the simplified job shop environment used in the learning factory training. Lastly, the training is split into phases that teach specific groups of competencies. In conclusion, the learning results can then be evaluated in regards to the set teaching objectives, which are presented in the next section.

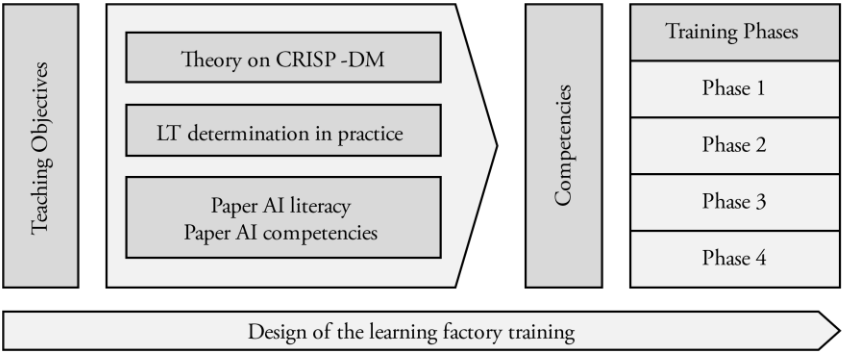

Figure 1: Methodology on the design of the learning factory training

5. Teaching objectives

According to the Learning Factory Curriculum Guide by Tisch et al. (2013), the first step in creating a new learning factory training is the determination of teaching objectives. Overall, the new training aims on teaching participants the necessary competencies to setup a data mining process according to CRISP-DM using ML approaches. Regarding CRISP-DM, the didactic objectives are the secure handling of the process steps themselves as well as understanding the impact the process itself has on the overall quality of the ML prediction models. Another key objective regarding CRISP-DM is to teach the participants the overall relevance of specific process knowledge in order to develop precise ML models.

In the field of ML, the learning factory training contains four teaching objectives. The first one is to provide participants with a comprehensive overview of the technologies and regression methods underlying the concept (A). Building on this theoretical foundation, the second teaching objective of the learning factory training is to emphasize the importance of comparing available learning models with reference to each individual case in application (B). The third teaching objective is defined by the ability of the participants to understand the respective areas of application of ML and the actions required to use it effectively and efficiently. This implies the participants to the importance of the interaction of process knowledge, precise data mining and the selection of sufficient ML processing applications (C). The fourth teaching objective is to enable the participants to independently design a ML application following the CRISP-DM using ML regression models (D).

Since the teaching objectives focus on individual participants, single or multiple quantitative indicators cannot measure the teaching objectives. Therefore, the game phases need to be designed to only allow participants to proceed within the learning factory workshop, if they fully enabled themselves to make use of the competencies taught in the previous game phase. This implies a strong connection of the success of a game phase from the participant's perspective with the achievements of the defined learning objectives from the teacher's perspective.

6. Identification of required competencies

6.1. CRISP-DM

CRISP-DM is a widely used, industry-independent standard process for data mining, which is divided into the six stages "business understanding", "data understanding", "data preparation", "modeling", "evaluation" and "deployment" (Chapman et al. 2000). It has already been applied by Lingitz et al. to the specific problem of forecasting LT (Kristoffersen et al. 2019; Lingitz et al. 2018). As a target-oriented use of ML models on real-world problems is influenced by the decisions made within the complete process – from data generation to deploying a final model – it is useful to carry out a structured data mining process to make appropriate model assumptions (Awad/Khanna 2015; Kuhn/Johnson 2013). The first phase, business understanding, contains the definition of business and data mining goals. Additionally, it analyses the current situation and defines the desired outcome of the whole process (Chapman et al. 2000). In the data understanding phase initial data are being collected, described, and the overall data quality is evaluated (Chapman et al. 2000). The data preparation phase deals with the selection of adequate data for the modelling phase and prepares the data. This is done by cleaning, adding further data or integrating further data sources as well as formatting the data (Chapman et al. 2000). The preparation phase also includes the crucial task of selecting relevant data attributes (e.g. number of workstations per order), also called features, or the creation of more suitable features (Kuhn/Johnson 2013). During the modelling phase, curated regression methods are modelled, and a generated test design evaluates the models. In the evaluation phase, the results are evaluated and follow up steps are determined regarding a potential deployment. The evaluation either leads to adjustments of the previous phase and therefore to a repetition of the modelling process or to an approval of a sufficiently performing model. If the model gets approved, the deployment phase starts, containing planning the deployment and runtime management as well as assessing the project within a project report. In the context of deployment, regular model updates are required as characteristics of manufacturing systems or processes can change (Deep/Singh 2015; Hammami et al. 2017). Because of its recursive structure, CRISP-DM underlies a continuous improvement process and thus serves a deployment more than a snapshot as a final output.

6.2. Lead time determination supported by machine learning

In practice, traditional methods for the determination of LT often react insufficiently to changing environmental influences, especially in complex production environments like job shop manufacturing. Possible environmental influences are e.g. machine breakdown or fluctuations of the production load (Ludwig/Nyhuis

1992; Wildemann et al. 2005). Traditional approaches for LT determination include methods based on estimates, historical values, logistic models and simulation (Ziarnetzky/Mönch 2016). The limitations of most of the classical methods can be explained by too simplified assumptions. Approaches such as flow degree oriented scheduling or simulation aim to overcome these limitations. However, these do not lead to a significant improvement of the planning quality in relation to the increased effort (Lödding 2016; Ludwig/Nyhuis 1992; Nyhuis/Wiendahl 2012).

According to Ludwig/ Nyhuis (1992), LT prediction contains order characterization, the processing sequence, current work in progress and the capacity situation within the production. In complex production environments, LT are influenced by many input factors. However, employees in production planning are often not fully aware of the relevant input factors and their effects (Halevi 2010; Schuh et al. 2019). To counter this, companies can utilize production data, which is often available in large quantities (Chen et al. 2014).The increased variety and quantity of data in companies can be used to evaluate data and derive insights for the planning process (Awad/Khanna 2015; Buxmann/Schmidt 2019; Ertel 2013; Gentsch 2018). However, the numerous factors influencing LT in job shop manufacturing make it difficult to master the planning process, even if all relevant data is recorded. This underpins the potential of ML in LT determination as already proven in several case studies (Gyulai et al. 2018; Lingitz et al. 2018). Table 1 presents an overview of used data, regression methods and evaluation methods of case studies in complex production environments.

All studies listed use regression methods for the ML-based determination of LT and use either simulation or real data. Overall, information such as order data, data on the system status, material or employee related data is included (Burggraf et al. 2020; Gyulai et al. 2018; Welsing et al. 2021). Since the importance of different input factors varies in the context of the respective company, industry-specific input factors are not further discussed in this chapter.

Table 1 also shows the regression methods selected in each case. Used models where regression trees (Öztürk et al. 2006; Schuh et al. 2019), support vector machines (Alenezi et al. 2008), multilayer neural networks (Wang/Jiang 2019), or linear regression methods (Sabuncuoglu/Comlekci 2002). Depending on the case study, multilayer neural networks (Asadzadeh et al. 2011; Kramer et al. 2020; Wang/Jiang 2019), the random forest model (Gyulai et al. 2018; Lingitz et al. 2018), or supported vector machines (Alenezi et al. 2008) exhibited the best prediction performance. Overall, simple regression models such as the decision tree or linear regression show poorer forecasting performance compared to more complex models (Asadzadeh et al. 2011; Gyulai et al. 2018; Kramer et al. 2020; Lingitz et al. 2018). The evaluation of the results in the analyzed studies are based on the deviations between planned and actual values. The evaluations are carried out with different evaluation ratios, whereby Root Mean Square Error (RMSE) or the normalized form NRMSE and Mean Absolute Percentage Error (MAPE) are frequently applied.

	Data		Methods									Evaluation		
	simulated dataset	real dataset	neural network	multilayer neural network	k-nearest-neighbor	supported vector machines	random forest	decision trees	boosted decision trees	linear regression	others	RMSE or NRMSE	MAPE	others
Alenezi et al. (2008)	x			x		x					x	x	x	
Asadzadeh et al. (2011)		x		x						x	x		x	
Gyulai et al. (2018)		x				x	x	x		x	x	x		
Kramer et al. (2020)		x		x			x	x	x	x		x		
Lingitz et al. (2018)		x	x		x	x	x	x	x	x	x	x	x	x
Öztürk et al. (2006)	x							x		x	x			x
Sabuncuoglu/Comlekci (2002)	x									x	x			x
Schuh et al. (2019)*		x						x						x
Wang/Jiang (2019)**		x	x	x								x		x
Welsing et al. (2021)		x				x	x	x	x		x			x

* prediction of transition time. ** prediction of order finish date

Table 1: ML methods used to determine LT in existing literature

Since the study on LT determination using different ML methods by Kramer et al. (2020) originates from the Leuphana University of Lüneburg, it serves as the basis for one phase of the learning factory training. The study examines a dataset of a real job shop production for an investigation period of about one year. On average, the initial method to determine LT (operation-specific execution time plus a generalized transition time) deviated by -12.17 days with a standard deviation of 12.95 days from the actual LT. Henceforward, a clear need for an improved determination method exists. Kramer et al. show with their comparison of the RMSE values from the different regression models that all chosen models outperformed the original procedure and that multilayer neural networks presented the best overall accuracy. They conclude, that the average deviation between actual and planned LT of the test dataset had decreased to 0.43 days with a standard deviation of 7.98 days using a multilayer neural network (Kramer et al. 2020). As the study shows sufficient results determining LT using ML, the case can function as a useful case for the learning factory training. Thus, the learning factory training uses an anonymized version of the analyzed dataset for one of the training phases. The absolved training enables participants to apply the gained competencies on a real-world dataset.

6.3. Competencies on determining lead times supported by CRISP-DM using ML

For the documentation and description of competencies relevant to the professional field, the matrix structure developed by Abel et al. for the assessment of professional-methodical competencies is used (Abel et al. 2013). It subdivides a main competence (e.g. ability to perform a method) into several sub-competencies. Each sub-competence is assigned the respective action, e.g. analysis of the actual process. Furthermore, required professional knowledge and underlying conceptual knowledge is assigned to each sub-competence.

To master real-world ML-related tasks in production environments, owning specific competencies is crucial. The necessary competencies to accomplish LT determination derive from three major fields. The fields are A) combine ML related theory with production process knowledge, B) handling CRISP-DM related processes and C) setting up a ML application. Field A reflects the main sub-competencies identified for AI literacy by Long/Magerko (2020). As the paper addresses AI applications overall, the 16 identified competencies were reduced to ten by using pairwise comparison against the relevance for a production environment.

Competence	Sub-Competence	ID
Handle ML theory in the context of production	Distinguish between general and narrow AI	1
	Recognize that computers perceive the world using sensors	2
	Recognize how computers make decisions.	3
	Understand that agents are programmable	4
	Compare different learning approaches	5
	Justify an algorithm from mathematical/statistical perspective	6
	Compare ML tools to each other empirically	7
Prepare a dataset for ML	Identify features that make an entity “intelligent”	8
	Understand that data requires process knowledge and interpretation	9
	Define scope of data mining project	10
	Setup data mining project	11
	Collect initial data	12
	Describe, examine and check the data quality	13
	Choose the relevant data for the modelling phase	14
	Format the dataset	15
Setup a ML model	Choose sufficient regression method	16
	Select applicable coding tools	17
	Avoid of the effects of overfitting	18
	Provide appropriate performance metric for algorithms	19
	Assess ML performance for the overall problem	20
	Conduct review process	21

Deploy a ML model	Determine steps needed prior to deployment	22
	Plan runtime management	23
	Implement ML into existing processes for deployment	24
Apply ML on new production related problems	Differ between challenging and suitable problems for ML	25
	Be aware of problems related to data bias	26
	Compare differences in interpretability of learned models	27
	Evaluate the effects of ML decisions	28
	Imagine possible future applications of ML	29

Table 2: Necessary competencies to develop a ML application following CRISP-DM

The 13 sub-competencies regarding CRISP-DM (field B) were extracted directly from the original method description (Chapman et al. 2000) and experiences of the authors applying the process as described in the fourth section. The original six phases of CRISP-DM have been consolidated into three major competencies being data preparation, model setup as well as model deployment. Field C consists of the twelve sub-competencies Danyluk/Buck identified to apply AI from a data science perspective (Danyluk/Buck 2019). After removing duplicate entries, the catalog is reduced from 35 to 29 sub-competencies overall. The sub-competencies originate from five identified competencies as shown in Table 2.

7. Learning factory training on lead time determination using CRISP-DM and machine learning

To enable sufficient knowledge transfer in a learning factory training, it is necessary to bring the competencies to be taught into a logical sequence so that game events are understandable to the participants. Therefore, in the present case, it is necessary to integrate the need for LT prediction into the game phases. According to the Learning Factory Curriculum Guide published by Tisch et al. (2013), the first didactic transformation defines the production type, purpose and target group of the learning factory training. To demonstrate the capabilities of ML methods applied on LT determination in a complex environment, for the new training a job shop scenario similar to the production system analyzed by Kramer et al. (2020) is applied to the learning factory. The purpose of the workshop is to enable participants to develop a ML application following CRISP-DM. Adapted from the original Leuphana Learning Factory, the target group consists mainly of professionals in PPC positions as well as undergraduate students and consultants. As the competencies to be taught have been identified in the last section, the second didactic transformation follows. It contains the definition of manufacturing processes, the manufactured product, the intended learning process as well as associated teaching methods (Tisch et al. 2013). Based on existing structures, the job shop of the learning factory handles the maintenance and repair of pre-assembled damaged model cars. The intended learning process, illustrated in Table 3, first provides learners with the necessary process knowledge for the production system. Second, learners record and prepare the dataset for ML modelling using conservative calculative methods. As a third phase, participants set up and deploy the ML model. In the final phase, learners apply a ML model on a real-world dataset.

#	Training Phase	Content	Desired result	Competen-cies
1	Socialize with the production system	Participants gain process knowledge by performing repair jobs on model cars using predefined work plans. One player schedules production, estimating LT.	LT do not meet estimations.	9
2	Record and prepare the dataset for ML modelling	Utilization of digitalization technologies to record data needed for LT determination. LT prediction using traditional tabular calculation software.	LT determination improves, but still lacks in precision.	1, 2, 3, 9, 10, 11, 12, 13
3	Set up and deploy the ML model	Participants prepare the recorded data for processing using ML. Application of regression models using RapidMiner. Evaluation on test data and another game round.	Determined LT are significantly more precise than in phase 2.	4, 5, 6, 7, 8, 9, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 28
4	Apply ML model on real-world dataset	Application of the gained competencies from phases one and two on an anonymized real-world dataset.	Participants archive results similar to those by Kramer et al.	4, 5, 6, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29

Table 3: Designed phases for the learning factory training

In the first phase, participants get in touch with the production process. In preparation on the first game round, participants receive a brief introduction to the key production logistics figures and performance indicators. Their task is to carry out defined maintenance orders on vehicles. Work plans and pre-picked part sets are provided for this purpose. Individual participants work at individual workstations, responsible for only one area, for example front wing repair. One player is given the task of transforming customer orders into production orders. The correct scheduling of different desired delivery dates is achieved by determining the planned LT based on the work plans and the current capacity utilization of the production areas. The latter must first be estimated in a self-managed manner for example by vision. The planned and actual LT of the production orders are tabulated by the instructors. In the course of production, the participants then discover that unforeseen waiting and non-productive times occur for some production orders. Reasons for delays are several predefined circumstances within the production that force delays. Circumstances are missing material at workstations or bottleneck workstations creating queues. Additionally, a high amount of work in progress and orders running through a high number of workstations lead to volatile LT that are hard to determine manually. At the end of the round, a discussion of the recorded LT takes place in the plenary. Phase 1 enables participants to gain necessary process knowledge to extract relevant data in the upcoming rounds, therefore addressing learning objectives B and C.

The second training phase starts with a theoretical input on CRISP-DM. The task of the participants is then to carry out the structured recording of the data relevant for LT calculation. For this purpose, the data needed must first be determined and then recorded. Various RFID and barcode-based time recording systems are available for this purpose. Subsequently, a second round of the game is carried out. Next, the recorded dataset is viewed by the participants and its quality is evaluated using standard office software. The second phase concludes with a theoretical input on ML methods and regression models. Additionally, the software tool RapidMiner (Mierswa/Klinkenberg 2021) is introduced. Due to its intuitive user interface and well-structured procedure for testing ML models, RapidMiner is used for the development of the ML application (Usuga Cadavid et al. 2020). Parallel to the theory input, one of the instructors statistically expands the recorded dataset. That way a dataset containing approx. 10,000 entries is generated from the initial 30 production orders carried out within the game round. Within the second game phase, participants gain competencies regarding recording relevant data and the importance of process knowledge to select suffient measurement points in the process. Phase 2 therefore addresses teaching objectives A and C.

The third phase addresses the preparation of the dataset, the modeling of the ML application and the evaluation of predicted LT using test data as well as a new game round. Again, participants first receive a theory input and then work on the task independently. First, participants evaluate the prediction quality of the estimation method used in the second round of the game and then answer some given questions about the data set using either tabular calculation or RapidMiner. The answers to the questions serve as a basis to perform the feature extraction. Subsequently, different regression methods are run in RapidMiner and the results are discussed in plenary. The third phase ends with another game round played, in which LT the best suited model is deployed into the manufacturing execution system by RapidMiner and then prediction for new arriving orders are made based on the trained ML model. A final discussion of the results takes place before the last phase starts. To master the third game phase, participants need to handle a full data mining process, containing gaining process knowledge, recording relevant data and developing a suitable ML model. To get sufficient results, teaching objectives A, B and C need to be achieved. To further check the teached competencies, game phase four confronts the participants with a new production process deriving from a real-world scenario.

In the fourth phase, the participants set up a ML supported LT determination application on a real-world case. The dataset provided to the participants is a anonymized version of the dataset of a job shop production used by Kramer et al. (2020). As Kramer et al. outline, the job shop is structured in 14 shop sections with three to six workstations assigned to each section. The dataset contains nine main production routes, containing changing work contents and job loops. The recorded maintenance orders vary in terms of external factors influencing the LT. Factors such as environmental conditions or the capacity status of the whole production system influence the LT as well as the amount of work needed to repair or maintain the given unit. This concludes in different workstation schedules for each order, which results in varying durations in process as well as transition times (Kramer et al. 2020). After applying ML by using RapidMiner (Mierswa/Klinkenberg 2021), the results, which should align with the findings from Kramer et al. (2020), are discussed. The successful development of a ML supported application on LT determination based on a real-world dataset by the participants confirms the achievement of teaching objectives A, B and C. Furthermore, the successful completion of phase four also attests the achievement of learning objective D – therefore completing the learning factory workshop.

Figure 2 on the next page shows exemplary findings that the participants could make within training phase four. The graphs demonstrate that the original procedure mostly overestimates the actual LT. In contrast, the graph created by multilayer neural networks aligns significantly more sufficient to the actual LT of those 40 orders. The results also show that a correction of the original procedure based on only the average deviation improves the prediction performance, but still lacks in precision as observed at orders number 12, 23 and 31. The workshop closes with a group discussion on further PPC tasks that could be addressed using similar ML approaches.

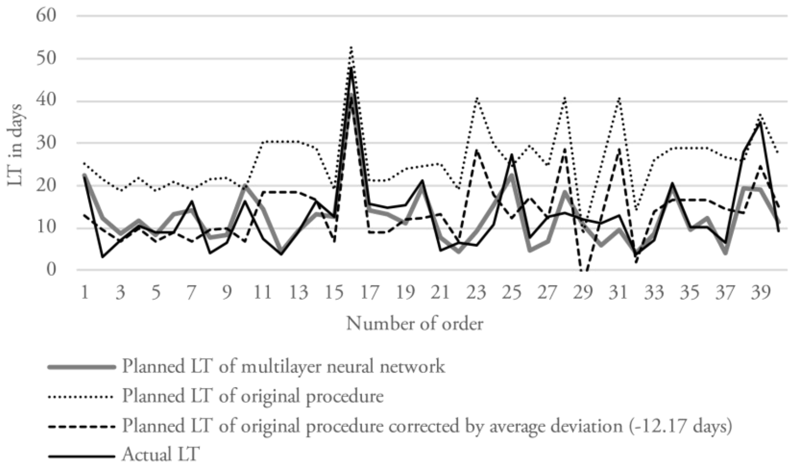

Figure 2: Excerpt of 40 orders of the test set for illustrating purposes

Figure 3 sums up how the training process enables participants to develop new ML based applications. Phase one allows the participants to gain process knowledge necessary to setup a targeted data mining process. Phases two teaches participants competencies needed to setup and execute CRISP-DM. Phase three uses the recorded dataset to teach learners the necessary competencies on developing and deploying a ML application. The fourth and final phase confronts the participants with a dataset from a real-world job shop production. Participants need to transfer the gained competencies on the given dataset to prove their ability to develop new ML applications on production related problems.

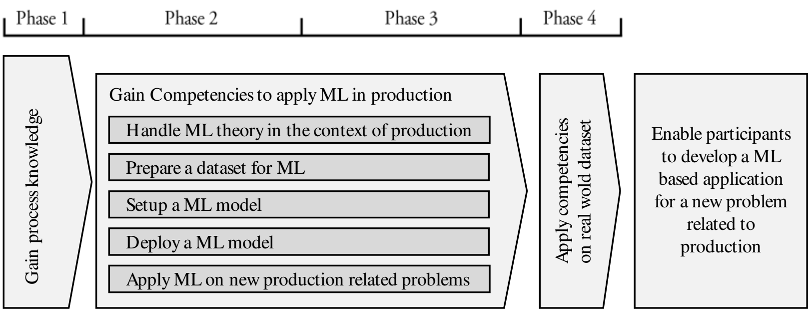

Figure 3: The learning factory training enables participants to develop new ML based applications

8. Conclusions and Outlook

Developing ML supported applications for PPC tasks requires certain competencies. Learning factories can help to acquire these competencies without influencing live production operations. For this purpose, the respective learning factories must be precisely tailored to the topics taught to enable the participants to gain the necessary competencies. Existing learning factories address ML focusing on tasks such as predictive maintenance, quality control or activity and position recognition (Daniyan et al. 2020; Hofmann et al. 2020; Oberc et al. 2020). To teach participants the necessary competencies to setup a data mining process according to CRISP-DM using ML, a multi-phase learning factory training is required. The process described in this chapter represents an approach to teach essential topics in the context of ML in production environments. Determining LT in a job shop environment serves as an example for participants to explore the topics of data mining, regression modelling and data interpretation. Follow-up tasks are to evaluate the learning success and iteratively improve the learning factory training. Additionally, the concept could be enhanced by integrating different PPC tasks or learning methods like reinforcement learning.

Acknowledgements

Funded by the Lower Saxony Ministry of Science and Culture under grant number ZN3489 within the Lower Saxony "Vorab" of the Volkswagen Foundation and supported by the Center for Digital Innovations (ZDIN).

References

Abel, M., Czajkowski, S., Faatz, L., Metternich, J., Tenberg, R. (2013). Kompetenzorientiertes Curriculum für Lernfabriken. *Werkstattstechnik Online : Wt, Springer VDI Verlag, Düsseldorf*, *103*(3), 240–245.

Abele, E., Chryssolouris, G., Sihn, W., Metternich, J., ElMaraghy, H., Seliger, G. et al. (2017). Learning factories for future oriented research and education in manufacturing. *CIRP Annals*, *66*(2), 803–826.

Abele, E., Metternich, J., Tisch, M. (2019). *Learning Factories*. Cham: Springer International Publishing.

Abele, E., Metternich, J., Tisch, M., Chryssolouris, G., Sihn, W., ElMaraghy, H. et al. (2015). Learning Factories for Research, Education, and Training. *Procedia CIRP*, *32*, 1–6.

Alenezi, A., Moses, S. A., Trafalis, T. B. (2008). Real-time prediction of order flowtimes using support vector regression. *Computers & Operations Research*, *35*(11), 3489–3503.

Asadzadeh, S. M., Azadeh, A., Ziaeifar, A. (2011). A Neuro-Fuzzy-Regression Algorithm for Improved Prediction of Manufacturing Lead Time with Machine Breakdowns. *Concurrent Engineering*, *19*(4), 269–281.

Awad, M., Khanna, R. (2015). *Efficient learning machines: Theories, concepts, and applications for engineers and system designers. The expert's voice in machine learning.* New York, NY: ApressOpen.

Burggraf, P., Wagner, J., Koke, B., Steinberg, F. (2020). Approaches for the Prediction of Lead Times in an Engineer to Order Environment—A Systematic Review. *IEEE Access*, *8*, 142434–142445.

Buxmann, P., Schmidt, H. (2019). Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens. In P. Buxmann & H. Schmidt (Eds.), *Künstliche Intelligenz* (pp. 3–19). Berlin, Heidelberg: Springer Berlin Heidelberg.

Cachay, J., Wennemer, J., Abele, E., Tenberg, R. (2012). Study on Action-Oriented Learning with a Learning Factory Approach. *Procedia - Social and Behavioral Sciences*, *55*, 1144–1153.

Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer (2000). CRISP-DM 1.0 Step-by-step data mining guide.

Chen, M., Mao, S., Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, *19*(2), 171–209.

Chollet, F. (2018). *Deep learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek* (K. Lorenzen, Trans.). Frechen: mitp.

Daniyan, I., Mpofu, K., Oyesola, M., Ramatsetse, B., Adeodu, A. (2020). Artificial intelligence for predictive maintenance in the railcar learning factories. *Procedia Manufacturing*, *45*, 13–18.

Danyluk, A., Buck, S. (2019). Artificial Intelligence Competencies for Data Science Undergraduate Curricula. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 9746–9747.

Deep, K., Singh, P. K. (2015). Design of robust cellular manufacturing system for dynamic part population considering multiple processing routes using genetic algorithm. *Journal of Manufacturing Systems*, *35*, 155–163.

Deslauriers, L., Schelew, E., Wieman, C. (2011). Improved Learning in a Large-Enrollment Physics Class. *Science*, *332*(6031), 862–864.

Döbel, Leis, Molina Vogelsang, Petzka, Riemer, Rüping et al. (2018). *Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung*. München.

Ertel, W. (2013). *Grundkurs Künstliche Intelligenz*. Wiesbaden: Springer Fachmedien Wiesbaden.

Gentsch, P. (2018). *Künstliche Intelligenz für Sales, Marketing und Service.* Wiesbaden: Springer Fachmedien Wiesbaden.

Gyulai, D., Pfeiffer, A., Nick, G., Gallina, V., Sihn, W., Monostori, L. (2018). Lead time prediction in a flow-shop environment with analytical and machine learning approaches. *IFAC-PapersOnLine*, *51*(11), 1029–1034.

Haghighi, A., Zadeh, N. S., Sivard, G., Lundholm, T., Eriksson, Y. (2014). *Digital Learning Factories: Conceptualization, Review and Discussion*, Göteborg.

Halevi, G. (2010). *Handbook of production management methods*. Oxford: Butterworth-Heinemann.

Hammami, Z., Mouelhi, W., Ben Said, L. (2017). On-line self-adaptive framework for tailoring a neural-agent learning model addressing dynamic real-time scheduling problems. *Journal of Manufacturing Systems*, *45*, 97–108.

Hofmann, C., Patschkowski, C., Haefner, B., Lanza, G. (2020). Machine Learning Based Activity Recognition To Identify Wasteful Activities In Production. *Procedia Manufacturing*, *45*, 171–176.

Keepers, M., Romero, D., Hauge, J. B., Wuest, T. (2020). Gamification of Operational Tasks in Manufacturing. In B. Lalic, V. Majstorovic, U. Marjanovic, G. von Cieminski, & D. Romero (Eds.), *IFIP Advances in Information and Communication Technology. Advances in Production Management Systems. The Path to Digital Transformation and Innovation of Production Management Systems* (Vol. 591, pp. 107–114). Cham: Springer International Publishing.

Kramer, Wagner, Schmidt (2020). Machine Learning-Supported Planning of Lead Times in Job Shop Manufacturing. In B. Lalic, V. Majstorovic, U. Marjanovic, D. Romero, & G. von Cieminski (Eds.), *IFIP Advances in Information and Communication Technology: Vol. 591. Advances in Production Management Systems. The Path to Digital Transformation and Innovation of Production Management Systems: IFIP WG 5.7 International Conference, APMS 2020, Novi Sad, Serbia, August 30 - September 3, 2020, Proceedings, Part I* (1st ed.). Cham: Springer International Publishing; Imprint: Springer.

Kristoffersen, E., Aremu, O. O., Blomsma, F., Mikalef, P., Li, J. (2019). Exploring the Relationship Between Data Science and Circular Economy: An Enhanced CRISP-DM Process Model. In I. O. Pappas, P. Mikalef, Y. K. Dwivedi, L. Jaccheri, J. Krogstie, & M. Mäntymäki (Eds.), *Lecture Notes in Computer Science. Digital Transformation for a Sustainable Society in the 21st Century* (Vol. 11701, pp. 177–189). Cham: Springer International Publishing.

Kuhn, M., Johnson, K. (2013). *Applied Predictive Modeling*. New York, NY: Springer New York.

Lingitz, L., Gallina, V., Ansari, F., Gyulai, D., Pfeiffer, A., Sihn, W., Monostori, L. (2018). Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer. *Procedia CIRP*, *72*, 1051–1056.

Lödding, H. (2016). *Verfahren der Fertigungssteuerung: Grundlagen, Beschreibung, Konfiguration* (3. Auflage). *VDI-Buch*. Berlin, Heidelberg: Springer Vieweg.

Long, D., Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. In *Bernhaupt, Mueller et al. (Hg.) 2020a – Proceedings of the 2020 CHI* (pp. 1–16).

Ludwig, E., Nyhuis, P. (1992). Verbesserung der Termineinhaltung in komplexen Fertigungsbereichen durch einen neuen Ansatz zur Plan-Durchlaufzeit-Ermittlung. In W. Brauer, W. Görke, H. Rininsland, & M. Syrbe (Eds.), *Informatik aktuell. Information als Produktionsfaktor* (pp. 473–483). Berlin, Heidelberg: Springer Berlin Heidelberg.

Martinez, P., Vargas-Martinez, A., Roman-Flores, A., Ahmad, R. (2020). A science mapping study on learning factories research. *Procedia Manufacturing*, *45*, 84–89.

Mierswa/Klinkenberg (2021). *RapidMiner Studio.*

Nyhuis, P., Wiendahl, H.-P. (2006). Logistic Production Operating Curves – Basic Model of the Theory of Logistic Operating Curves. *CIRP Annals*, *55*(1), 441–444.

Nyhuis, P., Wiendahl, H.-P. (2012). *Logistische Kennlinien: Grundlagen, Werkzeuge und Anwendungen* (3. Aufl. 2012). *VDI-Buch*. Berlin, Heidelberg: Springer.

Oberc, H., Fahle, S., Prinz, C., Kuhlenkötter, B. (2020). A Practical Training Approach in Learning Factories to Make Artificial Intelligence Tangible. *Procedia CIRP*, *93*, 467–472.

Öztürk, A., Kayalıgil, S., Özdemirel, N. E. (2006). Manufacturing lead time estimation using data mining. *European Journal of Operational Research*, *173*(2), 683–700.

Rammer, C., Bertschek, I., Schuck, B., Demary, V., Goecke, H. (2020). *Einsatz von Künstlicher Intelligenz in der Deutschen Wirtschaft: Stand der KI-Nutzung im Jahr 2019*. ZEW-Gutachten und Forschungsberichte. Berlin.

Sabuncuoglu, I., Comlekci, A. (2002). Operation-based flowtime estimation in a dynamic job shop. *Omega*, *30*(6), 423–442.

Sackey, S. M., Bester, A., Adams, D. (2017). INDUSTRY 4.0 LEARNING FACTORY DIDACTIC DESIGN PARAMETERS FOR INDUSTRIAL ENGINEERING EDUCATION IN SOUTH AFRICA. *South African Journal of Industrial Engineering*, *28*(1).

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 210–229.

Schäfers, P., Schmidt, M. (2015). Entwicklung eines integrativen Logistikmodells für die unternehmensinterne Lieferkette. *ZWF Zeitschrift Für Wirtschaftlichen Fabrikbetrieb*, *110*(12), 775–778.

Schmidt, M., Maier, J., Grothkopp, M. (2020). Eine bibliometrische Analyse - Produktionsplanung und -steuerung und maschinelles Lernen (wt Werkstattstechnik online 4-2020).

Schuh, G., Gützlaff, A., Cremer, S., Schopen, M. (2020). Understanding Process Mining for Data-Driven Optimization of Order Processing. *Procedia Manufacturing*, *45*, 417–422.

Schuh, G., Prote, J.-P., Sauermann, F., Franzkoch, B. (2019). Databased prediction of order-specific transition times. *CIRP Annals*, *68*(1), 467–470.

Tisch, M., Hertle, C., Cachay, J., Abele, E., Metternich, J., Tenberg, R. (2013). A Systematic Approach on Developing Action-oriented, Competency-based Learning Factories. *Procedia CIRP*, *7*, 580–585.

Usuga Cadavid, J. P., Lamouri, S., Grabot, B., Pellerin, R., Fortin, A. (2020). Machine learning applied in production planning and control: a state-of-the-art in the era of industry 4.0. *Journal of Intelligent Manufacturing*, *31*(6), 1531–1558.

Veza, I., Gjeldum, N., Mladineo, M., Celar, S., Peko, I., Cotic, M. et al. (2017). Development of Assembly Systems in Lean Learning Factory at the University of Split. *Procedia Manufacturing*, *9*, 49–56.

Wang, C., Jiang, P. (2019). Deep neural networks based order completion time prediction by using real-time job shop RFID data. *Journal of Intelligent Manufacturing*, *30*(3), 1303–1318.

Welsing, M., Maetschke, J., Thomas, K., Gützlaff, A., Schuh, G., Meusert, S. (2021). Combining Process Mining and Machine Learning for Lead Time Prediction in High Variance Processes. In B.-A. Behrens, A. Brosius, W. Hintze, S. Ihlenfeldt, & J. P. Wulfsberg (Eds.), *Lecture Notes in Production Engineering. Production at the leading edge of technology* (pp. 528–537). Berlin, Heidelberg: Springer Berlin Heidelberg.

Wiendahl, H.-P. (2014). *Betriebsorganisation für Ingenieure* (8., überarbeitete Auflage). München: Hanser.

Wildemann, H., Zäh, M. F., Müller, N., Krauß, U., Loth, M. (2005). Wandlungsfähige Auftragsabwicklung als Voraussetzung für effizientes Produzieren in Netzwerken. In O. K. Ferstl, E. J. Sinz, S. Eckert, & T. Isselhorst (Eds.), *Wirtschaftsinformatik 2005* (pp. 83–101). Heidelberg: Physica-Verlag HD.

Yang, H., Lu, W. F., Lin, A. C. (1992). Intelligent Process Planning Using a Machine Learning Approach. *IFAC Proceedings Volumes*, *25*(28), 147–151.

Zhang, W., Cai, W., Min, J., Fleischer, J., Ehrmann, C., Prinz, C., Kreimeier, D. (2020). 5G and AI Technology Application in the AMTC Learning Factory. *Procedia Manufacturing*, *45*, 66–71.

Ziarnetzky, T., Mönch, L. (2016). Incorporating engineering process improvement activities into production planning formulations using a large-scale wafer fab model. *International Journal of Production Research*, *54*(21), 6416–6435.