Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 119 (2023) 514–520

www.elsevier.com/locate/procedia

33rd CIRP Design Conference

Identification and prediction of standard times in machining for precision steel tubes through the usage of data analytics

Julian Weller[a,*], Daniel Roesmann[b], Sönke Eggert[c], Sebastian von Enzberg[a], Iris Gräßler[b], Roman Dumitrescu [a,d]

[a]Fraunhofer Institute for Mechatronic Systems Design, Zukunftsmeile 1, 33102 Paderborn, Germany
[b]Chair for Product Creation, Heinz Nixdorf Institute, Fürstenallee 11, 33102 Paderborn, Germany
[c]BENTELER Steel/Tube, Residenzstraße 1, 33104 Paderborn, Germany
[d]Chair for Advanced Systems Engineering, Heinz Nixdorf Institute, Fürstenallee 11, 33102 Paderborn, Germany

* Corresponding author. Tel.: +49-5251-5465213 ; *E-mail address:* julian.weller@iem.fraunhofer.de

Abstract

An approach to predict standard times for upcoming individual customer orders based on historical production orders is developed. The approach follows the CRISP-DM model with regard to the steps data preparation, modeling and evaluation. A precision steel tube straightening machine is provided as a validation example. Furthermore, the paper focuses on how the standard time prediction approach can be successfully integrated into the IT infrastructure of a production focused company. Especially data availability and accessibility in enterprise resource planning systems are discussed. The presented approach offers users the value of more precise scheduling in production than established approaches.

Peer review under the responsibility of the scientific committee of the 33rd CIRP Design Conference

Keywords: Standard Times, Data Analytics, Regression, ERP Integration

1. Motivation

Manufacturing companies are increasingly operating in markets characterized by customized products and shorter product and technology life cycles. Constantly changing market conditions cause companies to adapt quickly to new conditions. Scheduling of machining processes relies on the prediction of production times. Within mass customization, these times vary for each product due to individual machine parameters. Crucial to lean production and adaptability is the accuracy of time data of production systems. These depend on the production capabilities as well as the product-specific characteristics. The standard time as a subset of the time data contributes to precise production and capacity planning as well as preliminary order costing. A high degree of accuracy is necessary to avoid large deviations in production planning and preliminary costing. Current methods of determining time data are reaching their limits due to the difficulties resulting from high volume and variety of data. Within production, information and communication technologies, such as manufacturing execution systems and production data acquisition systems, are used to increase transparency and to improve productivity. These possibilities are still being used to a limited degree in industry to determine standard times.

The research goal of this paper is to present a procedure model to develop a method to predict standard times for upcoming individual customer orders based on historical production orders. It is supported by a case study in the precision steel tube manufacturing company *BENTELER*. For this purpose, different data analytics methods are used. Furthermore, the integration of the prediction into the IT-Infrastructure of a manufacturing company is discussed. The results enable companies to further develop a prescriptive analytics approach in production planning.

2212-8271 © 2023 The Authors. Published by Elsevier B.V.

Peer review under the responsibility of the scientific committee of the 33rd CIRP Design Conference
10.1016/j.procir.2023.01.011

2. Methodology

A methodological approach according to Ulrich [1] is chosen which enables the comprehensive development and initial method for identification and prediction of standard times in machining. This was chosen because of the practice-oriented and integrative character and the reference to engineering as well as management sciences. The development of the approach is based on actual practical problems in industry. The analysis of the problem in chapter 3 provides the basis for this paper. Based on literature research, deficits in the current state of the art are discussed. In chapter 4, specific preliminary work from the field of scheduling and time management are analyzed with regard to the objectives. For this purpose, the identified deficits are compared with the preliminary work. As an argumentative-deductive analysis in the spectrum of methods, a systematic literature search is conducted according to Briner & Denyer [2]. The reviews of existing approaches result in a concrete scientific need for action. Chapter 5 presents the method developed based on the identified deficits and need for action. The procedure is based on the CRISP-DM [3] model with a focus on the steps data preparation, modeling and evaluation. The procedure model is derived from the combination of data analytics and standard time specific methods. Based on an expert interview, data of the ERP system and machine specific data serve as an initial database. The developed method is validated in an industrial evaluation with the production system in Schloss Neuhaus of the company *BENTELER Steel/Tube* in chapter 6. For this purpose, a prototypical implementation is carried out with the use of Jupyter notebooks. Within the study, the procedure and the use of different algorithms are considered. Chapter 7 summarizes the results and findings of the work and draws a conclusion regarding the achievement of the objectives. The chapter provides advice for practice and discusses possible further developments and research needs based on the results.

3. Problem analysis

According to Deuse et al., time management is an elementary function in manufacturing companies that frames production planning [4]. Time management is divided into the steps of determination, structuring and provision with continuous administration. Within production planning, it serves to schedule orders and allocate resources, equipment and employees [5]. Only time management ensures that decisions such as capacity planning, cost calculation, offer calculation and remuneration calculation are secured [6]. Kuhlang et al. [7] and Petzelt [8] emphasize that the analysis and improvement of production processes is an essential task for modern time management.

Within time management, different time data exist. Petzelt [8] divides time data into actual time, normal time, target time and standard time: Actual times represent the time required in reality for the production process. Normal times are times standardized to the performance level of 100 %. Target times are based on the analysis of actual times and/or normal times, considering influencing factors. The standard times result from the addition of surcharges due to distribution time (time required due to additional activities to the work task) and setup time.

The standard times are of particular relevance for planning and scheduling. Standard times form the basis for planning production processes, such as predicting and scheduling future orders and determining bottlenecks. A decisive property of standard times is their reproducibility. Reproducibility includes the traceability and the repeatability at different times and static requirements [9]. According to lean production, an inaccurate estimation of production time leads to various types of waste in production. The waste occurs in the form of waiting times, overproduction and rejects. Therefore, the forecast quality of the time data is essential for subsequent planning and optimization steps. Standard times are used in various IT systems for production planning. Exemplary systems are enterprise resource planning (ERP) systems, manufacturing execution systems (MES) and advanced planning and scheduling (APS) systems [10, 11].

According to Deuse et al., Burggraf et al. and Kuhlang et al. there are different methods for determining time data [4, 6, 7]. On the one hand, there are methods for recording actual times and on the other hand for determining target times. *Actual times* can be recorded by video recordings, multi-moment frequency distributions, self-recording/questioning, automated time registration, REFA time recording, multi-moment time measurement methods, among others. *Target times* can be determined in the following ways: Time recording with performance level, planned time determination, system of predetermined times, time module catalogue, comparison and estimation, estimation with time classes, process time formulas, nomographs, calculation in case of uncertainty and simulation tools. Due to the rise of product and production process complexity, existing methods are reaching their limits. The high number of product variants implies that a lot of data has to be collected, processed and stored [12]. As the machinery of a company is often diverse, especially with regard to manufacturer, year of construction and dimensions, the time data must be instantiated specifically for every machine. This leads to a high effort (time and cost) with the existing methods. Furthermore, within industrial practice, it can be seen that different time data are applied in different business units and IT systems. These time data are often based on individual, implicit expert knowledge of the business units [13, 14].

The practical example of precision steel tubes illustrates a limitation for interlinked production systems. The relevance of certain master data changes during the production process due to physical changes of the tubes. Furthermore, the prediction models for machines vary depending on production methods. The different models require specific input. Examples are a changed diameter, length or tensile strength of the steel. Standard times are specific to a machine or a grouped unit of machines (one process step). For precision steel tubes, they are usually expressed in seconds per defined amount of product.

4. Related work

Current research on time management [15–18] recommends the use of data analytics methods. Eraslan emphasizes that the main advantage of these methods, apart from their cost structure, is the flexibility of their application. A respective adaptation to the application case must take place [15].

In addition to the conventional methods of time management described in chapter 3, it is also possible to determine standard times using data analytics methods. Compared to conventional methods, data analytics methods often require less manual effort. To integrate these advanced methods into a process flow, the time data must satisfy defined statistical requirements [4]. This includes accuracy, which describes the difference between the acquired and real time data [5]. A particular challenge is to map the changes to the machines over their lifetime into the digital model. Continuous optimization and change must be considered in the data analytics methods [19]. Otherwise, the underlying data quickly becomes obsolete and loses accuracy for the application.

Jodlbauer et al. propose the integration of these into one model through a supplemented parameterization. The goal is to use existing historical data and keep the data gathering process as lean as possible [9]. Ramirez et al. use linear regression methods [18] and Kutschenreiter et al. implement neural networks to determine the production time of similar products on similar production lines [16]. Gelmereanu et al. recommend the use of neural networks to determine default values for cycle times in milling processes [17].

All presented approaches deal with specific sub-problems and do not generalize their approach sufficiently. Solutions regarding data governance and integration are missing. Hence, no method with a holistic approach could be identified to our knowledge.

5. Prediction of standard times in machining

To fill this research gap, within this paper a new method is developed. The method is used to create a model that predicts standard times for upcoming individual customer orders based on historical production orders. The developed method consists of **five general logical process steps.** They are supported by twelve independent **work packages** to generate new data of historical data for the **standard time determination** (see Fig. 1 and Fig. 3**)**. The process steps represent the technical framework, whereas the work packages provide users with a methodological structure. The procedure is based on the CRISP-DM approach.

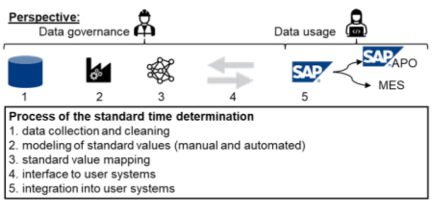

Fig. 1. Different views on the process of standard time determination

First, the five general logical process steps are presented (see Fig. 1): The **input data are collected and cleaned** (1). The input data is used for **modeling** and **mapping** the chosen approach (2,3) [20]. Integration into a user system (e.g. ERP) is ensured by an interface (4).

In preparation, it is important to include all stakeholders' interests into the modeling of the standard time. For example, production planners want to utilize their machines to the maximum degree and meet their key performance indicators (KPI) such as overall equipment effectiveness (OEE). The sales team wants to be able to calculate appealing prices. Hence, it is very important to create a transparent process to match partially contradicting interests.

The first step is to select the *input data*. Supported by a literature review (see section 2) a solution space is constructed in a morphological box (see Fig. 2). The columns represent the different process steps. The rows represent solution elements for the steps. A suitable combination needs to be found with expert knowledge to ensure interoperability between the solution elements. As a data source, every element of the automation pyramid can be used. Aggregated and structured data is usually preferred. The input from non-machine-readable data sources should not be neglected. Proprietary tools (like Excel sheets) and expert knowledge often contain crucial knowledge about the process. For *modeling*, common data science tools can be used. Established standard time determination techniques complete the solution space. The resulting model display (*mapping*) depends on the used algorithm. ERP systems usually represent the target system for the standard times. Hence, the usable *interface* can be derived from the system in use. Depending on the chosen solution, everything between a manual update up to an automated update pipeline can be applied. The modeling algorithm has a direct impact on how easy it can be integrated into systems with a relational database. If not organized externally, the model outputs might need to be transformed into a table format for easier integration.

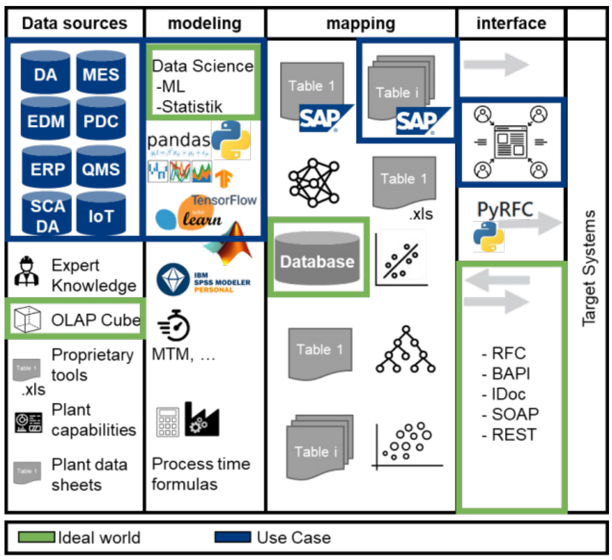

Fig. 2. Morphological box of the solution space

The proposed five general logical process steps are scalable and can be applied to small, medium and large enterprises. To support the application within companies, twelve independent supporting work packages were derived based on the process steps. These are each supported by suitable methods. This enables easy application as well as distribution of responsibilities within the company. An extract of the developed method is visualized in the BPMN-like Fig. 3.

First, the reference system is described in a **use case**. The use case needs to be understood with expert knowledge (**process understanding**). This is described with the help of a use case diagram in an expert interview. Subsequently use case specific sets of **solution elements** are generated. The decision is driven by the company specific boundaries. Next on, the **steps data collection, data cleaning and data preparation** are performed until a sufficient data quality is reached (support, see Fig. 5). The sufficiency of the data highly depends on the underlying sub systems. Next, the **data** is **prepared, analyzed and modeled**. These process steps can be automated and standardized to a high degree. A **manual replacement** process is proposed for processes with an insufficient amount or quality of data. After a successful evaluation of the model, it is released and transferred to live.

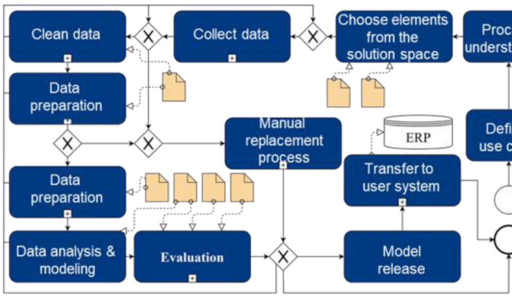

Fig. 3. Standard time renewal workflow

6. Application and validation

At *BENTELER Steel/Tube*, the area with steel tube straightening machines is most crucial. This is due to both the percentage deviation between actual and planned production time and the associated cost impacts. When producing a seamless precision steel tube, the product is manufactured in a sequence of different steps. Each step ensures the quality demanded by the customer. To supply the specified geometrically correct dimensions, one needs to straighten steel tubes after heat treatment. Part of the straightening machine is an in-line measurement and quality inspection process. Before delivery, the tubes are cut into the specified length and packaged according to specifications. Input like the exact production sequence and setup times would improve the accuracy of standard times. But during the planning process it is not possible to generate a definite schedule because of disrupting impacts of the production. These impacts lead to deviations from planning. Thus, the given production planning approach can not include these variables to calculate standard times. This limits the possible accuracy in the planning process.

First, the decisions on data inputs and on the tool to be used are made based on the use case. The solution space is determined with the help of the morphological box from chapter 4. Within an expert interview it was determined that the data from the ERP system and the MES system of the company will be used. Python is used to implement the data analysis approach. Within the analysis, different algorithms are used, which also require different modeling. A data pipeline for the preprocessing of the input data for standard time values based on the library pandas and scikit learn is suggested. Different regression algorithms are compared. Applied regression algorithms are linear regression, random forest, MLP neural network, decision tree and k-means. The R-squared score is used to compare the algorithms. All regressors are optimized by grid search and compared to the existing manually generated standard times.

Modeling with the given infrastructure follows the CRISP-DM model. Iterative steps in the development process are part of the process and ensure a growth in model maturity.

6.1. Data collection and data cleaning

First, checks regarding data quality and sufficiency for the use case are carried out. Less than 5000 products were produced in the given time frame. One needs to ensure the adequate distribution of the data before splitting them into training, test and validation data. The data is sorted regarding product characteristics. Numeric data is divided into three percentiles and categorial data into different variables. Then a random pick chooses the same amount of every product that exists more than the average of all products. Like this, data smoothing is applied without losing the products that only exist in a smaller amount. Example parameters are production time (Y), physical dimensions of the tube, material characteristics, machine parameters and job-related master data.

As a first check, the deviation of standard times is calculated for the given dataset. The deviation between reality and plan is high with a very low R squared score (below 0,35). This will be used as a benchmark for the model in development. After ensuring the validity of the data, the data with basic graphical analysis are explored. Therefore, process experts were involved to give further input for other possible (mechanical engineering related) parameters and their recombination. In this phase heatmaps and a pairwise scatterplots are used (see Fig. 4). No singular correlation could be identified to use one dimensional linear approach. Thus, a combination of different parameters must be used to improve the model.

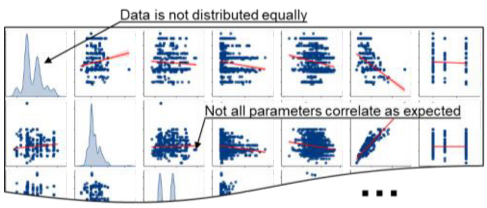

Fig. 4. Explorative data analysis of historical production data

The pipeline is then executed as shown in Fig. 5. The data is split into training and test data. A cross validation is used to evaluate the model. The first iteration resulted in a lack of sufficient parameters. Thus, additional master data and material data were acquired and added. Once more, the hypothesis that a standard time prediction model needs to be related to the type of machine is confirmed. Also, data is needed from different levels of the automation pyramid.

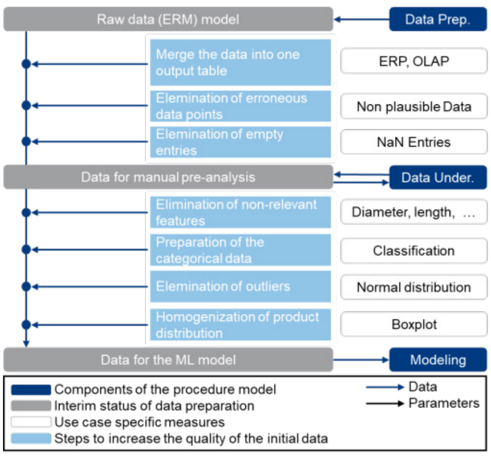

Fig. 5. Workflow to preprocess the data

6.2. Modeling

A supervised learning approach was chosen (input and result known). The problem (predict task) suggests the usage of regressors. Classification algorithms could be used to group products into non continuous scales for their standard time. The regressors are chosen via recommendations from other papers [15, 16, 18] and a review of different regressors. The applied regressors are the following:

- (Multiple) linear regression, Ridge regression
- Nearest neighbor regression
- Decision tree regressor, Random Forest regressor
- MLP regressor (neural network with LBFGS stochastic gradient decent)

The input data is defined by different tube related material and production master product data (ERP-System). The new standard time is predicted with the usage of the real production time as the target value. The prepared data is split into training and test data. A cross validation is used. The pipeline consists of a numerical scaler, a categorical encoder and the chosen ML algorithm. Every algorithm was optimized by grid search in the given grid of hyperparameters. The grid was defined by using all possible combinations or by a restriction of the solution space through expert knowledge. After finding a set of hyperparameters for each algorithm, the algorithms are compared. The best algorithm will be deployed after evaluation and comparison with the existing solution. The structure of the modeling approach is visualized in Fig. 6.

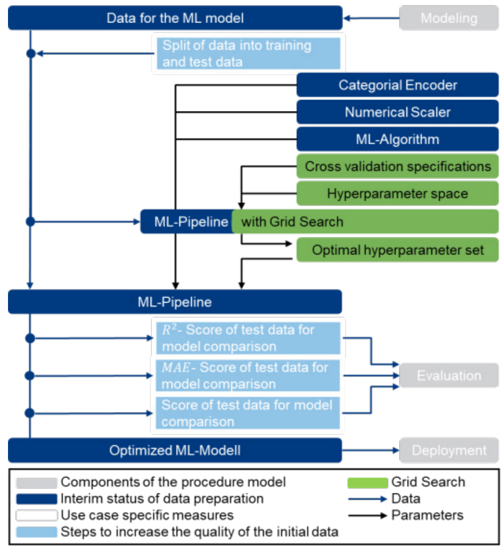

Fig. 6. ML pipeline to optimize the standard time model via grid search

6.3. Evaluation and deployment

As a metric to compare the different algorithms, the R squared score is used. The common ML score ensures that the models can be validated against each other and against an optimum. Three standard production products were compared in parallel to ensure that important products are modelled well. The resulting R squared score from the chosen straightening machine data is displayed in Fig. 7. All algorithms resulted in a model that is better than the current implementation. The current state is defined by the comparison between existing table like standard time and real production time. This suggests an optimization potential of the data input. The upstream hyperparameter optimization already produces the best possible results from the models with the given initial data. Best scores are achieved by the advanced algorithms The relative noise in the data rises with decreasing production time. Batches which ran way less than half a shift usually give data where the prediction has the strongest deviation. Prediction for bigger batches works best.

Fig. 7. Different model R^2 scores on the same data split

The ML model is developed in JupyterLab and versioned in GitLab. The integration of the result depends heavily on the used model. Regressors can be easily implemented into an ERP system. The relational database structure of ERP systems demands a different approach for neural networks, decision trees and nearest neighbor. They need to be coded into the system or connected via an interface.

The ERP system can request thousands of standard times for a planning forecast. Hence, the model needs to provide a very minimal cycle time. Another approach is to integrate the data as part of the manufacturing order in the ERP System. A model can also be transformed into a table with fixed values to integrate it into ERP. This approach has less accuracy and makes products out of the provided range unpredictable.

7. Conclusion and outlook

The presented approach enables a production driven company to improve their accuracy in production planning. The improved accuracy has effects on the overall logistical and production related efficiency of the given organizational unit.

A method for structuring the introduction of new standard times is presented. The method proved to be applicable in a medium sized enterprise. Generalized steps help to structure the company specific process and create a solution space. Depending on the digitalization maturity and data flow concepts of a company, a specific solution can be chosen. All provided tools help to structure the process of creating a company specific standard time renewal. A validation in a different sized company should be conducted in the future work to further improve its general applicability.

One of the key findings of the case study for the definition of standard times is the lack of visibility of derivations between planned and actual resulting values. Hence, a new KPI is needed to track the company's performance in its own planning process. A common database can further reduce manual time input into the model development. The analysis showed that a machine specific dependency for the optimization of the ML algorithms cannot be avoided. Thus, a generalized algorithm to produce standard times cannot be provided. Additional research needs to be invested to find the optimal integration technique into modern day ERP Systems.

Future work can be invested into standardizing the developed solution into a building brick for an analytics platform. Predictive approaches build the foundation for prescriptive use cases. One of the main challenges remains in the dependency of expert knowledge and data availability of the needed machine or product data. Machine Learning algorithms proved to be the most suitable data driven way of generating new and accurate standard times. The provided work shows that there is a high potential in the application of data analytics on standard times in machining.

Acknowledgements

This research was funded by the German Federal Ministry of Education and Research (BMBF) in the project VIP4PAPS, grant number 03VP10031. The contents of this publication are the sole responsibility of the authors.

References

[1] H. Ulrich, "Die Betriebswirtschaftslehre als anwendungsorientierte Sozialwissenschaft," in *Die Führung des Betriebes*, Sandig zu seinem 80. Geburstag gewidmet, Curt, Ed., 1981.

[2] R. B. Briner and D. Denyer, *Systematic Review and Evidence Synthesis as a Practice and Scholarship Tool*, 2012.

[3] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," *Journal of Data Warehousing*, vol. 5, no. 4, pp. 13–22, 2000.

[4] J. Deuse, N. West, and M. Syberg, "Rediscovering scientific management. The evolution from industrial engineering to industrial data science," *Int. J. Prod. Manag. Eng.*, vol. 10, no. 1, pp. 1–12, 2022, doi: 10.4995/ijpme.2022.16617.

[5] B. Lotter and H.-P. Wiendahl, *Montage in der industriellen Produktion*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[6] P. Burggraf, J. Wagner, and B. Koke, "Artificial intelligence in production management: A review of the current state of affairs and research trends in academia," in *2018 International Conference on Information Management and Processing (ICIMP)*, London, 2018, pp. 82–88.

[7] P. Kuhlang, S. Hempen, W. Sihn, and J. Deuse, "Systematic improvement of value streams - fundamentals of value stream oriented process management," *IJPQM*, vol. 12, no. 1, p. 1, 2013, doi: 10.1504/IJPQM.2013.054860.

[8] D. Petzelt, *Konzept zur rechnerunterstützten Bestimmung von Sollzeiten in der Montageplanung*. Zugl.: Dortmund, Techn. Univ., Diss., 2010. Aachen: Shaker, 2010.

[9] H. Jodlbauer, K. Palmetshofer, and S. Reitner, "Implicit Determination of Standard Times," *Wirtschaftsinf*, vol. 47, no. 2, pp. 101–108, 2005, doi: 10.1007/BF03250982.

[10] I. Gräßler and A. Pöhler, "Produktentstehung im Zeitalter von Industrie 4.0," in *Springer eBook Collection, Handbuch Gestaltung digitaler und vernetzter Arbeitswelten*, G. W. Maier, G. Engels, and E. Steffen, Eds., Berlin: Springer, 2019, pp. 1–21.

[11] V. Stich, D. Oedekoven, and T. Brosze, "Informationssysteme für das Logistikmanagement," in *Logistikmanagement*, G. Schuh and V. Stich, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 257–304.

[12] H. Jodlbauer, "Discussion of the standard processing time," *International Journal of Production Research*, vol. 42, no. 7, pp. 1471–1479, 2004, doi: 10.1080/00207540310001645129.

[13] I. Gräßler, A. Pöhler, and J. Hentze, "Decoupling of Product and Production Development in Flexible Production Environments," *Procedia CIRP*, vol. 60, pp. 548–553, 2017, doi: 10.1016/j.procir.2017.01.040.

[14] P. Gönnheimer, M. Netzer, L. Mohr, G. von Hörsten, and J. Fleischer, "Erhöhung der Skalierbarkeit von KI-Anwendungen in Produktionsanlagen," *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, vol. 115, 7-8, pp. 517–519, 2020, doi: 10.3139/104.112318.

[15] E. Eraslan, "The Estimation of Product Standard Time by Artificial Neural Networks in the Molding Industry," *Mathematical Problems in Engineering*, vol. 2009, pp. 1–12, 2009, doi: 10.1155/2009/527452.

[16] I. Kutschenreiter-Praszkiewicz, "Application of artificial neural network for determination of standard time in machining," *J Intell Manuf*, vol. 19, no. 2, pp. 233–240, 2008, doi: 10.1007/s10845-008-0076-6.

[17] C. Gelmereanu, L. Morar, and S. Bogdan, "Productivity and Cycle Time Prediction Using Artificial Neural Network," *Procedia Economics and Finance*, vol. 15, pp. 1563–1569, 2014, doi: 10.1016/S2212-5671(14)00626-1.

[18] J. Ramirez, R. Guaman, E. C. Morles, and L. Siguenza-Guzman, "Prediction of Standard Times in Assembly Lines Using Least Squares in Multivariable Linear Models," in *Communications in Computer and Information Science, Applied Technologies*, M. Botto-Tobar, M. Zambrano Vizuete, P. Torres-Carrión, S. Montes León, G. Pizarro Vásquez, and B. Durakovic, Eds., Cham: Springer International Publishing, 2020, pp. 455–466.

[19] R. Joppen, S. von Enzberg, D.-I. Arno Kühn, and D.-I. Roman Dumitrescu, "A practical Framework for the Optimization of Production Management Processes," *Procedia Manufacturing*, vol. 33, pp. 406–413, 2019, doi: 10.1016/j.promfg.2019.04.050.

[20] A. Kühn, R. Joppen, F. Reinhart, D. Röltgen, S. von Enzberg, and R. Dumitrescu, "Analytics Canvas – A Framework for the Design and Specification of Data Analytics Projects," *Procedia CIRP*, vol. 70, pp. 162–167, 2018, doi: 10.1016/j.procir.2018.02.031.