Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 81 (2019) 874–879

www.elsevier.com/locate/procedia

52nd CIRP Conference on Manufacturing Systems

Data Mining Definitions and Applications for the Management of Production Complexity

Günther Schuh[a], Gunther Reinhart[b], Jan-Philipp Prote[a], Frederick Sauermann[a], Julia Horsthofer[b]*, Florian Oppolzer[a], Dino Knoll[b]

[a]*Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, 52074 Aachen, Germany*
[b]*Institute for Machine Tools and Industrial Management (iwb), Technical University of Munich, 85748 Garching, Germany*

* Corresponding author. Tel.: +49 89 28915576; fax: +49 89 28915576. *E-mail address:* julia.horsthofer@iwb.mw.tum.de

Abstract

Production complexity has increased considerably in recent years due to increasing customer requirements for individual products. At the same time, continuous digitization has led to the recording of extensive, granular production data. Research claims that using production data in data mining methods can lead to managing production complexity effectively. However, manufacturing companies widely do not use such data mining methods. In order to support manufacturing companies in utilizing data mining, this paper presents both a literature review on definitions of data mining, artificial intelligence and machine learning as well as a categorization of existing approaches of applying data mining to manage production complexity.

Keywords: data mining; machine learning; artificial intelligence; production complexity

1. Introduction

In the last decade, manufacturing companies have faced an increasing demand for customization of products that could formerly be produced in mass production. This external trend has led to an increased internal production complexity [1]. In parallel, a trend of shop floor smartification through a roll-out of sensors, which increase communication between machines and employees, could be observed - a trend commonly known as *Industrie 4.0*. Hence, using large volumes of the generated production data in order to gain knowledge of relationships and interdependencies has become a research area of particular interest [2]. It is the enrichment of and pre-processing from raw to smart data that creates the actual value of Industrie 4.0. By usage of artificial intelligence (AI), more specifically machine learning (ML) and data mining (DM), data can be transformed into knowledge for various applications. As a result, it is expected that production managers will be able to master the aforementioned arisen production complexity effectively [3].

Accompanying challenges can be divided into two classes: (1) theoretical and (2) application-related challenges. The first challenge arises due to the strong linkage of the three major disciplines in this research area AI, ML and DM. So far, the relations and distinction between these terms have not been defined consistently within literature [4]. However, it is necessary to thoroughly understand the different terms in order to systematically evaluate concepts and methods for managing production complexity. The application-related challenge arises with the difficulty of implementing AI, ML and DM in practice due to various reasons such as low data integrity in databases [5] and difficult modeling of production knowledge [6].

10.1016/j.procir.2019.03.217

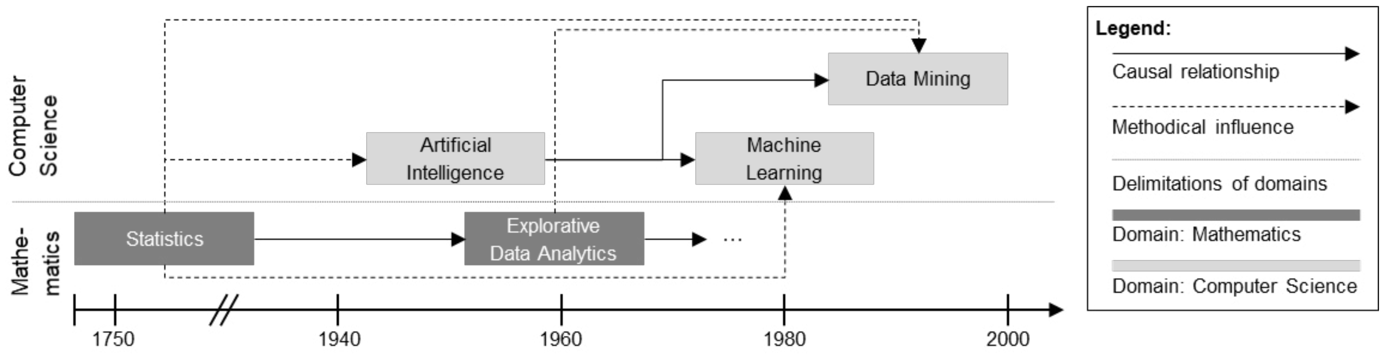

Fig. 1. Historical development of AI, ML and DM.

The goals of this paper are therefore (1) to create a common understanding for the terms DM, ML, AI and statistics from an application point of view (i.e. production), and (2) to support production managers to identify relevant use-cases for managing production complexity through DM.

The remainder of the paper is structured as follows. In section 2, the terms of AI, ML and DM are historically analyzed, defined in an application-oriented manner and subsequently separated from one another. Section 3 lays out a framework for managing production complexity with presenting and classifying existing applications of DM methods in manufacturing companies. Section 4 summarizes the results and gives an outlook on future research.

2. Definition of AI, ML and DM

2.1. Historical developments of the terms DM, ML and AI

The term AI refers to the eponymous field of science, which emerged under the influence of computer science, mathematics, neuroscience and other scientific disciplines and was shaped in by several phases of strong research activity and economic interest (cf. Fig. 1). From this, three basic objectives of these individual phases can be identified in retrospect [7, 8].

Objective 1: Develop a toolbox for imitating human thinking and actions

Gödel, Church and Turing, among others, laid the foundations of computer science and logic for computer technology in the 1930s. Programmable computers became available and, subsequently, the idea of automating human thinking and behavior arose [9]. In 1950, Turing described a theoretical concept, which later became known as the Turing Test, defining the branches and tools that would later be subsumed under the term *AI* [10, 11]. In the following years, after the goal of the so-called symbolic AI had been set, researchers were concerned with the hypothesis that intelligence on a human level could be achieved by modelling a sufficient amount of knowledge in form of logical connections and automated reasoning by computers [12]. These expert systems showed limitations of the symbolic AI approach that could not live up to expectations [8, 9].

Objective 2: Develop tools for solving specific problems

The pitfalls of symbolic AI stated above, led to a paradigm shift. Instead of modelling explicit knowledge, in 1987 Rumelhart and McClelland shifted the focus to the assumption that a computer can learn rules by observing connections in data, which moved especially the subject area of ML into focus [13, 14]. This insight enabled a movement summarized under the term connectionism. By linking many simple computing units in form of neural networks, a flexible yet robust architecture is created, countering the symbolic AI approach [12]. In addition to neural networks, other ML methods such as kernel methods (e.g. support vector machines), hierarchical and ensemble learning methods (e.g. decision trees) also gained acceptance [15]. In 2006, more extensive neural networks were introduced and deemed particularly useful for central problems of AI, especially with regard to vision and language. This field is better known as deep learning [4, 15].

Objective 3: Develop tools for identifying and explaining patterns in data

Since the late 1990s an inherent need for tools interpreting the vast and exponentially growing amounts of data stored in databases has emerged [9, 16]. Thus, the field of DM has developed from the environment of AI and under the influence of statistics, employing ML methods and statistical data analysis with the aim of addressing this need [9, 17]. In addition to gaining knowledge from data through DM, extensive end-to-end concepts have gradually developed, starting with company and task analysis through data acquisition and DM to the provision of software tools [16, 18].

2.2. Definitions

Based on the above introduced three objectives of the data science phases, the definitions of AI, ML and DM are derived.

AI seeks to enable computational agents to act and think rationally and intelligently [8, 11, 19]. The scientific goal of AI is to understand the principles of knowledge representation that enable intelligent behavior. The engineering goal of AI is to create computational agents that can solve real world problems as or more effectively and efficiently than humans [8, 11, 19]. The implementation of these premises has many different forms. Thus, AI can be seen as a toolbox whose subdomains deliver tools to create intelligent computational agents [8].

ML is a subdomain of AI and seeks to enable computational agents to gain task-related knowledge and solve task-specific problems. ML methods aim to optimize a performance criterion. This criterion acts as an indicator of the degree to which a given task is solved. ML methods enable computational agents to learn from (historical) data. The degree of solution of a task is optimized by learning. ML offers tools to AI and is thus the basis for further subdomains of AI. ML and statistics methods form the core of DM [8, 17, 20].

DM is another subdomain of AI and can be defined as a process that aims to generate knowledge from data and presents findings comprehensively to the user. Generating knowledge in the context of DM can be translated to the discovering of new and non-trivial patterns, relations and trends in data useful to the user. DM as a process involves, in essence, the collection and selection of data, the pre-processing of data, data analysis itself including the visualization of results, interpretation of findings, and the application of knowledge. To pre-process and analyze data, ML and statistics methods are deployed in DM. Findings from DM processes can be distinguished in descriptive ones, where knowledge is represented in form of models that depict patterns and relations in data and predictive ones, where knowledge is represented in a prediction of future conditions, trends and relations [16, 21, 22].

2.3. DM and its similarities and differences to AI, ML and statistics

While the term statistics hardly gets used synonymously for the other three terms, it is important to point out the scientific and methodical differences. For the comprehensiveness of this paper, similarities and differences between all of these four terms can be found in the matrix below (cf. Fig. 2)

As for the terms DM and AI, the difference lays in the relation. As DM is a subdomain of AI, it relates to the goals of AI but specifies and implements them. DM does not supply new methods to AI but employs methods that evolved in ML and statistics to extract knowledge from data [8].

The difference between ML and DM is that, while for both ML methods are used, they are used for different purposes and thus with different requirements. In ML, the knowledge is stored implicitly and serves the purpose of optimizing computational agents' performances. In DM, ML methods are employed so that knowledge is gained from data and is then stored and visualized explicitly, making it accessible and interpretable to the user [20].

Statistics supplies methods directly to DM. Statistics, as a subdomain of mathematics, is per definition a formal science. DM does not require the same formality, even when employing statistics methods. This allows DM to analyze data without hypotheses and is driven by results that are to be evaluated by experts, rather than precise reproducibility of the same. This less formal concept was introduced to the domain of statistics as Exploratory Data Analysis (EDA) before DM evolved, and has influenced the approach of DM towards statistics. The fewer formal requirements also enable DM to use statistical methods on data that has not been specifically designed for analyses [23]. This feature becomes especially important in the industrial context, where data integrity is still a big issue [5].

2.4. Insights and discussion

Throughout our research we found that the term DM had the least stringent definitions, whereas the definitions for ML were

	AI	**ML**	**DM**	**Statistics**
AI	AI seeks to enable computational agents to act and think rationally and intelligently [8, 11, 19].	ML is a subdomain of AI representing a practical implementation of AI premises by learning from data and experiences in order to optimize performance [8].	DM is a subdomain of AI that does not introduce further methods to the field of AI, but uses methods from AI (subdomains) to extract knowledge from data and make it accesible to the user [8].	Statistics is a formal science that alongside neuro sciences, philosophie and psychology influenced AI [8].
ML	ML is a subdomain of AI [8].	ML seeks to enable computational agents to gain task-related knowledge and solve task-specific problems [8, 17, 20].	While ML focuses on optimizing the performance of fulfilling a task, DM focuses on extracting knowledge and making it available to the user [20].	Statistics is a formalized subdomain of mathematics that seeks to find conclusions from data samples, while ML also applies statistical methods but seeks to fulfill a task efficiently [23].
DM	DM is a subdomain of AI [8].	ML methods, as well as statistical methods are applied in DM processes and form the core of DM [21].	DM is a process aiming to generate knowledge from data and presenting findings comprehensively to the user., Knowledge here means the discovery of new and non-trivial patterns, relations and trends in data useful to the user.[16, 21, 22].	DM neither necessitates a hypothesis nor the formality that statistics requires. DM applies statistical methods to data that has not intentionally been designed to be analyzed by statistical methods [23].
Statistics	Statistics has a strong methodical influence on AI and AI tools [8].	ML uses statistical methods [20].	DM uses stastical methods. Furthermore, the statistics subdomain EDA, has strongly influenced the concept of DM [23].	Statistics aims to gain quantitive information from observations in order to draw conlusions from the data (inductive) or describe the data (descripitve) [24].

Key: Definition Similarities Differences

Fig. 2. Definitions, similarities and differences of the terms AI, ML, DM and statistics.

rather congruent. The field of AI as such is so extensive that defining it as a toolbox serves the best structural framework for the definition of ML and DM

Overall, we deem a clear definition of the terms AI, ML and DM in the production management community as very important. A common understanding enables common progress through advancing and optimizing applications. Furthermore, the given definitions enable production managers to conduct a targeted research for applications based on the need for implicit or explicit knowledge.

3. Application of DM methods for managing production complexity

In order to support the identification of use-cases for the management of production complexity using DM, we outline relevant terms of production complexity and categorize existing approaches into six DM application categories (cf. Table 1). Thus, this paper enables the identification of relevant DM approaches regarding production complexity.

3.1. Production complexity

Production complexity can be distinguished in external and internal complexity. External complexity in production is created by increasing market and customer demands, as well as legal requirements. External complexity is reflected in a variety of products that are produced to meet market and customer demands. Internal complexity is caused by the implementation of product variety in an internal value chain.

To manage external complexity, product varieties have to be evaluated and, if possible, reduced. To reduce internal complexity, value chain processes must be optimized with regard to efficiency. [25]

3.2. Search and classification methodology

To find applications of DM suitable to support managing production complexity, we examined current classifications of DM applications in production or manufacturing literature. None of the eight publications found [3, 26-32] presented categories specific to production complexity. While explicit references of production complexity were missing, some of the presented categories related to production planning and control and decision support contained applications of interest to the topic of production complexity.

We then examined the applications in those categories and evaluated their contribution towards managing external or internal complexity. Additionally, we conducted a keyword-based research for applications of DM methods in managing production complexity outside the mentioned publications. Finally, we clustered the suitable applications found in 42 publications into six categories separately for internal or external production complexity. These categories have been developed at the interface of already existing DM applications in production management and the assumption that managing production complexity requires efficient support regarding product-based as well as process-based decisions in multi-variant value streams [33].

3.3. Applications

Our research showed that not all application categories were represented equally. 'Evaluating and preventing new variants' was the least represented category, with only two publications out of the 42 we found suitable for the topic of managing production complexity with DM application matching the category. 'Choosing dispatching rules and planning sequencing' as well as 'Process planning for new products/variants' were the two categories with the most findings (>10 publications each).

Table 1 represents the categories of DM applications suitable to manage production complexity that we identified in the applications and exemplary publications. To match the scope of this paper, we present the 16 most relevant and within each category diverse applications from publications.

Table 1. DM application categories and exemplary publications.

Complexity	DM application category	Exemplary publications
External	Evaluating and preventing new variants	Neis [34]
	Modulization and standardization	Agard, Kusiak [35], Romanowski, Nagi [36]
Internal	Process planning for new products/variants	Hochdörffer et al. [37], Denkena et al. [38], Wallis et al. [39]
	Choosing dispatching rules and planning sequencing	Bohnen et al. [40], Koonce, Tsai [41], Liu, Dong [42]
	Predicting and optimizing lead and cycle times	Cheng et al. [43], Gröger et al. [30], Backus et al. [44]
	Value stream complexity	Rozinat et al. [46], Park et al. [47], Lee et al. [48], Knoll et al. [49]

Evaluating and preventing new variants: Applications in this category aim to assess the costs and benefits of creating new variants to offer decision support.

Neis [34] presents an approach employing clustering methods to assess the costs of adopting new variants based off reference products. Products are initially clustered into product families and reference products are assigned based on the shortest distance to all other products within the cluster. Factoring in the distance between a new variant and the reference product, a cost function is calculated [34].

Modulization and standardization: Applications in this category aim to reduce product variety by identifying common parts and possible modules.

Agard and Kusiak [35] seek to identify subassemblies as modules using association analysis. Parts in customer orders are analyzed for common appearance in orders. Item sets (combinations of parts) that exceed the confidence levels are then examined regarding their feasibility as module [35].

Instead of modulization, Romanowski and Nagi [36] propose an approach for standardization based on the bill of material (BOM) using clustering to reduce product variety.

After products have been clustered into product families, the parts used for the products within a product family are clustered based on textual DM. The emerging cluster dendrogram is examined by a product expert for the possibility of standardized parts within a part cluster [36].

Process planning for new products/variants: In this category, applications aim to support planning processes for products and variants based on the existing portfolio.

Hochdörffer et al. [37] suggest using clustering methods to determine products requiring similar manufacturing technologies and similar capacity on these machines. The clusters can be used to determine processes for new variants or to optimize production networks [37].

Similarly, Denkena et al. [38] suggest clustering products based on processes but extend the approach by using k-nearest neighbor classification to classify new products/variants. Based on this classification processes from the nearest neighboring product/variant can be adopted or used as planning base [38].

Wallis et al. [39] propose using clustering and classification methods as well, but deploy them differently. Products are clustered first into part-based clusters and then, separately, into process-based clusters. Using the naïve Bayes function, process-based clusters are mapped onto part-based clusters. This allows more efficient assembly planning and exposes relations between variants and processes [39].

Choosing dispatching rules and planning sequencing: Applications in this category aim to support choosing dispatching rules based on the existing conditions and seek to make sequencing more efficient.

Bohnen et al. [40] present an approach for production levelling using clustering methods. Products are clustered based on their manufacturing requirements. Time blocks for production are then dedicated to the product families, which are sequenced based on the needed set-up change. This allows to efficiently minimize set-up costs and time [40].

Seeking to gain information about dispatching rules and factors influencing lead times Koonce and Tsai [41] propose employing decision trees. Initially, using evolutionary algorithms, for a realistic scenario dispatching rules are compared based on lead times. A decision tree is used to learn factors of different dispatching rules influencing lead time [41].

Following the goal of gaining information about dispatching rules and influencing factors, Liu and Dong [42] suggest an approach similar to Koonce and Tsai but suggest using artificial neural networks (ANN) that analyze and determine lead times of different dispatching rules [42].

Predicting and optimizing lead and cycle times: Applications in this category seek to help understand and assess factors that influence lead and cycle times.

Cheng et al. [43] propose an approach using decision trees to examine the influence of production staff on lead times and predict lead times based on staff set-up. Additionally, manufacturing tasks can be assigned based on individual performance. Correlations between staff set-up and lead times are analyzed via monitored manufacturing steps. [43].

A more general approach is presented by Gröger et al. [30], who suggest a combined approach of structured query language (SQL) and DM to quickly identify influence factors. Influence factors on key performance indicators (KPI), such as lead time, can be identified using data stored in structured data bases and accessed through SQL-queries. The data can then be directly analyzed without data preparation using classification methods that identify factors influencing the classifying KPI [30].

Backus et al. [44] propose using clustering methods and regression trees to predict cycle times for lots based on historical data. Previous lots are clustered based on common bottlenecks in the production system. Regression trees analysis is then used to determine factors influencing cycle time within common lots and thus enabling prediction [44].

Value stream complexity:

This category presents applications analyzing complexity from a process-oriented value stream perspective. Process mining (PM) has evolved as a DM method for discovering, analyzing and improving processes. PM extracts process models based on event data created during operations [45].

Rozinat et al. [46] applied PM to reduce the complexity of a wafer scanner testing process. Based on a discovery analysis, feedback loops and idle times were identified [46]. Park et al. [47] use PM to analyze a production process within the shipbuilding industry. When combining PM with data envelopment analysis (DEA), a variety of block types could be analyzed and differences between planned and actual operations were identified [47].

Within logistics we also identified PM applications across different industries (e.g. shipbuilding [48] and automotive industry [49]). The PM methods (mainly discovery [48] and conformance checking [49]) were combined with DM methods (e.g. clustering [48]) to improve processes. Exemplarily, Lee et al. [48] use PM and clustering to discover process models, iterations and bottleneck activities [48]. Contrastingly, Knoll et al. [49] address product and processes complexity using multidimensional PM to identify waste.

PM supports the value stream perspective both in production and logistics. Therefore, PM should be seen as a key DM technique for analyzing and reducing value stream complexity. Further research directions should address the integration of established lean principles for PM.

4. Findings and conclusion

At the interface of rising production complexity due to shifting market demands and vast amounts of production data, DM can be a valid tool to support managing complexity.

Most applications of DM in production management have so far been related to quality management. There are very few applications of DM directly related to production complexity. However, other applications of DM in other fields of production management serve the purpose of managing production complexity very well. We have presented some of these applications and plan to extend the categories in future work to present a holistic framework of DM, as well as other ML and AI applications able to cover all relevant aspects of managing production complexity.

This is a resupply of March 2023 as the template used in the publication of the original article contained errors. The content of the article has remained unaffected.

In order to verify the outlined framework, we plan to evaluate the methods by employing real process data. This is especially significant as many of the presented approaches have only been implemented using synthetic data.

A holistic framework and its validation in practice are thus the logical next steps.

References

[1] ElMaraghy, H; Schuh, G; ElMaraghy, W; Piller, F; Schönsleben, P; Tseng, M; Bernard, A. Product variety management. CIRP Annals 2013;62:629-652.

[2] Tao, F; Qi, Q; Liu, A; Kusiak, A. Data-driven smart manufacturing. Journal of Manufacturing Systems 2018;48:157–169.

[3] Choudhary, A; Harding, J; Tiwari, M. Data Mining in manufacturing - A review based on the kind of knowledge. Journal of Intelligent Manufacturing 2009;20:501-521.

[4] Goodfellow, I; Bengio, Y; Courville, A. Deep learning. Cambridge, Massachusetts, London, England: MIT Press, 2016.

[5] Schuh, G; Reuter, C; Prote, J; Brambring, F; Ays, J. Increasing data integrity for improving decision making in production planning and control. CIRP Annals 2017;66:425-8.

[6] Knoll, D; Prüglmeier, M; Reinhart, G. Predicting Future Inbound Logistics Processes Using Machine Learning. Procedia CIRP 2016;52:145-150.

[7] Chollet, F. Deep learning with Python. Shelter Island, NY: Manning (Safari Tech Books Online); 2018.

[8] Russell, S; Norvig, P. Artificial intelligence. A modern approach. 3rd ed. Upper Saddle River, NJ: Prentice-Hall; 2010.

[9] Bibel, W.; Ertel, W.; Kruse, R.: Grundkurs Künstliche Intelligenz. Eine praxisorientierte Einführung. 3rd ed.Wiesbaden: Springer; 2013.

[10] McCarthy, J; Minsky, M; Rochester, N; Shannon, CE. A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955. AI magazine 2006;27):12-4.

[11] Turing, AM. Computing Machinery and Intelligence. In: Mind 1950;59:433–460.

[12] Smolensky, P. Connectionist AI, symbolic AI, and the brain. Artificial Intelligence Review 1987;2:95-109.

[13] Shalev-Shwartz, S; Ben-David, S. Understanding machine learning. From theory to algorithms, Cambridge: Cambridge University Press; 2014.

[14] Rumelhart, D.; McClelland, J. Parallel distributed processing. Explorations in the microstructure of cognition (Series: Computational models of cognition and perception). 3. print Ed. Cambridge, Mass.: MIT Pr, 1987

[15] LeCun, Y; Bengio, Y; Hinton, G. Deep learning. Nature 2015;521:436-444.

[16] Fayyad, U; Piatetsky-Shapiro, G; Smyth, P. From data mining to knowledge discovery in databases. AI magazine 1996;17:37-54.

[17] Alpaydin, E. Introduction to machine learning. Cambridge, Massachusetts, London, England: MIT Press; 2009.

[18] Wirth, R; Hipp, J. CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th International Conference on the Practical Application of Knowledge Discovery and Data Mining 2000:29-39.

[19] Winston, PH. Artificial Intelligence. Boston: Addison-Wesley; 1993.

[20] Mannila, H. Data mining: machine learning, statistics, and databases. Proceedings of 8th International Conference on Scientific and Statistical Data Base Management 1996;8:2-9.

[21] Witten, IH; Pal, CJ; Frank, E; Hall, MA. Data mining. Practical machine learning tools and techniques. Cambridge: Morgan Kaufmann; 2017.

[22] Chapman, P; Clinton, J; Kerber, R; Khabaza, T; Reinartz, T; Shearer, C; Wirth, R. CRISP-DM 1.0. Step-by-step data mining guide. SPSS Inc.; 2000.

[23] Hand, DJ. Data Mining: Statistics and More? In: The American Statistician 1998;52:112-8.

[24] Dietrich, E; Schulze, A. Statistische Verfahren zur Maschinen- und Prozessqualifikation. München: Carl Hanser Verlag; 2014.

[25] Schuh, G; Kampker, A, editors. Strategie und Management produzierender Unternehmen. Handbuch Produktion und Management 1. Berlin, Heidelberg: Springer-Verlag; 2011.

[26] Otte, R; Otte, V; Kaiser, V. Data Mining für die industrielle Praxis. München: Hanser; 2004.

[27] Lieber, D. Data Mining in der Qualitätslenkung am Beispiel Stabstahlproduktion. Herzogenrath: Shaker; 2018.

[28] Wallis, R. Data-Mining-basierte Erstellung von Montagearbeitsplänen in der Digitalen Fabrik. Herzogenrath: Shaker; 2016..

[29] Harding, JA; Shahbaz, M; Srinivas; Kusiak, A. Data Mining in Manufacturing: A Review. International Journal of Production Research 2006;128:969-976.

[30] Gröger, C; Niedermann, F; Mitschang, B. Data Mining-driven Manufacturing Process Optimization. Proceedings of the World Congress on Engineering 2012;1-7.

[31] Wang, K. Applying data mining to manufacturing: the nature and implications. In: Journal of Intelligent Manufacturing 2007;18.487-495.

[32] Chen, MC. Configuration of cellular manufacturing systems using association rule induction. International Journal of Production Research, 2003;41:381-395.

[33] Hooshmand, Y; Köhler, P; Korff-Krum, A. Komplexitätsbeherrschung und Transparenzerhöhung in der Einzelfertigung. ProduktDaten Journal, 2013;2:55-9.

[34] Neis, J. Analyse der Produktportfoliokomplexität unter Anwendung von Verfahren des Data Mining. Herzogenrath: Shaker, 2015.

[35] Agard, B; Kusiak, A. Data Mining for Subassembly Selection. In: IEEE Transaction on Pattern Analysis and Machine Intelligence 2004;126:628-631.

[36] Romanowski, CJ; Nagi, R. A Data Mining Approach to Forming Generic Bills of Materials in Support of Variant Design Activities. In; International Journal of Production Research 2004;4:316-347.

[37] Hochdörffer, J; Laule, C; Lanza, G. Product variety management using data-mining methods. Reducing planning complexity by applying clustering analysis on product portfolios. IEEE International Conference 2017:593-7.

[38] Denkena, B; Schmidt, J; Krüger, M. Data Mining Approach for Knowledge-based Process Planning. Procedia Technology 2014;15:406-415.

[39] Wallis, R; Erohin, O; Klinkenberg, R; Deuse, J; Stromberger, F. Data Mining-supported Generation of Assembly Process Plans. Procedia CIRP, 2014;23:178-183.

[40] Bohnen, F; Maschek, T; Deuse, J. Leveling of low volume and high mix production based on a group technology approach. CIRP Journal of Manufacturing Science and Technology 2011; 4:247-251.

[41] Koonce, DA.; Tsai, SC. Using data mining to find patterns in genetic algorithm solutions to a job ship schedule. Computers & Industrial Engineering 2000;38:361-374.

[42] Liu, H; Dong, J. Dispatching rule selection using artificial neural networks for dynamic planning and scheduling. Journal of Intelligent Manufacturing 1996;7:243-250.

[43] Cheng, YJ; Chen, MH; Cheng, FC; Cheng, YC.; Lin, YS.; Yang, CJ. Developing a decision support system (DSS) for a dental manufacturing production line based on data mining. IEEE International Conference 2018; 638-641.

[44] Backus, P; Janakiram, M; Mowzoon, S; Runger, GC; Bhargava, A. Factory Cycle-Time Prediction With a Data-Mining Approach. IEEE Transactions on Semiconductor Manufacturing 2006;19:252-258.

[45] Van der Aalst, WM. Process mining: data science in action. Berlin Heidelberg: Springer; 2016.

[46] Rozinat, A; de Jong, IS; Gunther, CW; van der Aalst, WM. Process mining applied to the test process of wafer scanners in ASML. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 2009;39:474-479.

[47] Park, J; Lee, D; Zhu, J. (2014). An integrated approach for ship block manufacturing process performance evaluation: Case from a Korean shipbuilding company. International Journal of Production Economics 2014;156:214-222.

[48] Lee, SK; Kim, B; Huh, M; Cho, S; Park, S; Lee, D. Mining transportation logs for understanding the after-assembly block manufacturing process in the shipbuilding industry. Expert Systems with Applications 2013;40:83-95.

[49] Knoll, D; Reinhart, G; Prüglmeier, M. Enabling value stream mapping for internal logistics using multidimensional process mining. Expert Systems with Applications 2019; 124:130-142.

This is a resupply of March 2023 as the template used in the publication of the original article contained errors. The content of the article has remained unaffected.