VSM 4.0: Application Potentials of Data Science

B.A. Lina Döring
Master studies Data Science, University of Applied Science Bielefeld, Germany
lina.doering@hsbi.de

M.Sc. Tim Wollert
Doctoral Center for Social, Health and Economic Sciences, Magdeburg-Stendal University of Applied Sciences, Germany

DOI: http://dx.doi.org/10.25673/116282

Abstract
Value Stream Management (VSM) is a business methodology that focuses on optimizing the flow of materials and information across an organization and beyond to deliver maximum value to customers. It originated from lean manufacturing principles but has been adapted and applied in various domains, such as logistics and administration. To address some weaknesses of the conventional methodology and utilize the potential of increasing digitalization in supply chains in general and companies in particular, various studies explored the targeted application of modern technologies. A common approach in recent studies is to use a digital representation of the value stream, which is dynamically adjusted through continuous processing of operational business data. As reasoned by several studies, the application of data-driven techniques on current data and historical data in the areas of VSM offers several benefits and opens new opportunities in production and logistics management, such as real-time monitoring, early warning-system, enhanced decision-making, predictive analytics, discrete simulation and further ones. Combining VSM with Data Science techniques presents a synergy for organizations aiming to enhance efficiency and maximize value delivery. VSM provides a structured approach to visualizing and optimizing value streams, while Data Science techniques provide the means to gather and analyze big amounts of business data for improving and decision-making. By the present paper, the combined application of VSM and Data Science is investigated, aiming at the provision of an operation framework, which links the various elements of both domains.

1. Introduction
The digitalization of VSM, mentioned as Smart VSM, VSM 4.0, and Dynamic VSM, is made a research subject in several studies, taking different perspectives on the methodology and improvement approaches into account, as shown by [1]. In this context, data is described as an essential key driver for the digital transformation of companies, especially regarding value-adding processes. In addition, the management of data is a potential source of modern waste in terms of information logistics [2]. Both aspects are considered in terms of the future viability of VSM [3], [4]. In various studies data-processing techniques, such as data mining [5], [6], process mining [7], [8], data-based decision-making [9], [10] real-time data processing and data analytics in general [11] and further ones are introduced and critically discussed in terms of application fields. Therefore, the huge amount of business data is utilized for an automated mapping [4], [12] of the actual value stream, waste identification and elimination [13], [14], simulations of measures and target value streams for process improvement [15], [16] and further fields of application. All these data-driven techniques are implicitly related to the research area of Data Sciences. Most studies are limited to the consideration of solutions, combining single aspects of VSM, such as mapping and analyzing with individual data tools, such as machine learning (ML) and data-based decision-making. But for now in research, the merge of VSM and Data Science in its entirety, linking the phases of VSM, concretely from the mapping of the value stream over analyzing, designing to implementing (sometimes also monitoring) with the phases of Data Science projects.

This is the result of a systematic literature review according to PRISMA [17]. Based on an initial review of studies in the database Google Scholar (https://scholar.google.de/), by the search string (“value stream mapping” OR "value stream management”) AND “data science” 528 records, published between 2020 and 2024 are identified. Applying the inclusion criteria English or German, availability and relevant content, the number of records is reduced to 34 ones. The expression “VSM” is explicitly not included, due to different meanings of the abbreviation, such as “vector space method” and its implicit inclusion in the terms “value stream management” and “value stream mapping”.

Data Science, a vast field of research, necessitates a more structured approach to project and process management to facilitate its integration with VSM. Data Mining, also known as Knowledge Discovery, describes the entire process from specifying to understanding a data related problem over analyzing it (e.g. for hidden patterns) to gain insights and find solution approaches.
It is pointed out, there is no sharp and clear definition and common understanding of the terminology. In this context, the analysis phase centered around modeling data to address future challenges (for example, predicting future trends based on historical patterns), is also referred to as "Data Mining”. From a procedural perspective, the Data Mining process represents a specific methodology to be employed for Data Science applications [18], [19]. The CRoss Industry Standard Process for Data Mining (CRISP-DM) offers a standardized methodology for navigating through Data Science projects, ranging from the initial understanding of the problem or business context to the deployment of an effective model [19], [20]. Therefore, the framework CRISP-DM underlies the research, presented in the paper at hand.

The result of the literature review leads to two central research questions, which are investigated in the framework of the study at hand.

- **RQ1**: How to merge VSM and Data Science by mapping the phases of VSM with the CRISP-DM?
- **RQ2**: How to utilize data processing techniques and tools in the framework of Data Science in combination with the VSM procedure?

2. Methods

As stated in the literature review, some Data Science related techniques were already proposed to apply in the different VSM activities as the amount of data to process during VSM rises. But there is no holistic framework bringing together VSM and the Data Science stack. The paper at hand aims at closing this gap. For VSM, the actions that need to be taken in the four phases are explained in 2.1. To address the requirements for integrating big data into a tangible application, it is essential to delineate a structured approach that encompasses the necessary steps and considerations for effective implementation. To do that, the CRISP-DM was chosen and is explained in 2.2. The two approaches from the domains are then mapped in 2.3. To provide an operational framework, a selection of technologies and algorithms to incorporate Data Science in VSM is provided in 2.4.

2.1. VSM Procedure

VSM represents a holistic management approach to mapping, analyzing, and designing end-to-end value chains from suppliers to customers. It deals with eliminating waste according to Lean Management principles, aiming at the minimization of the overall lead time. Conventionally, this methodology relies on manual tools such as paper, pencil, and stopwatch [21], making it static in volatile environments, as reasoned by [22]. Recent research focuses on digitalizing VSM to overcome several disadvantages and take advantage of the increasing amount of business data. Therefore, in this section the procedure is described based on a four-phase model with reference to data-driven approaches, such as [3], [23], [24].

Phase 1: Value Stream Mapping
Value Stream Mapping refers to recording the actual value stream [21]. In addition to capturing the information and material flows in logical sequence, VSM-specific metrics are recorded, which are then consolidated in their entirety in a value stream map. Referring to a digital model of the value stream map, data from business application systems, such as enterprise resource planning (ERP) and manufacturing execution system (MES) as well as from machines and plants is utilized for deriving the status quo.

Phase 2: Value Stream Analysis
Value Stream Analysis, which depending on the literature considered can also include the phase of Value Stream Mapping, e.g. [21], describes a process of examining and evaluating the value stream’s efficiency based on wastes such as overproduction, waiting time, unnecessary transportation, excess inventory, defects, and similar ones [21]. By the gathered metrics, the overall performance is measured, whereas root cause analyses are conducted on the identified wastes.

Phase 3: Value Stream Design
In the third phase, a target value stream is designed, eliminating the weaknesses and wastes, identified in the previous phase. In this context, tools and principles of Lean Management are applied, such as a leveled utilization of all

resources, reduction of stocks, optimization of setup processes by SMED (Single Minute Exchange of Die), flow orientation, customer-centric and further ones [21].

Phase 4: Value Stream Planning
Value Stream Planning refers to the iterative implementation of measures to achieve the target value stream and gradually reduce waste while increasing efficiency. [25], [26], [27].

2.2. CRISP-DM

This work aims at providing an operational framework that support the implementation of Data Science technologies and models in the context of VSM. The CRISP-DM is a common cycle to follow when implementing a data related model for a certain task. Although it is not new, Data Science is currently gaining a lot of publicity, but there is still no standard approach or model fixed in a German DIN norm. The German Federal Ministry for Economic Affairs and Climate Action supports the non-profit organization DIN with shaping this. They published the second edition of a “Normungsroadmap KI” - a road map to a potential DIN norm for Artificial Intelligence (AI) in 2022 [28].

The CRISP-DM represents a comprehensive methodology applicable to data mining projects. It serves as a structured framework designed to guide the execution of the development of applications leveraging (big) data for specific objectives within a corporate environment. This methodology has six distinct phases and incorporates options for iterative adjustments for refinement [20]. The procedure model is shown in Figure 1 and elucidated in this section.

Phase 1: Business Understanding
At the start of a Data Mining project, there is the need for general business understanding. The project team needs to determine the general business objectives. When the group is an internal team, ideally these are known by the team. But, especially when an external service provider is involved, the objective should be stated to all stakeholders in the project team. In addition to that, the current situation of the company should be assessed. For example, the available resources as well as potential risks the company is facing should be identified. Subsequently, the objectives of Data Mining, which the organization aims to achieve through the application of Data Science Technologies, must be clearly defined to facilitate the measurement of their success. Concluding this initial phase, the formulation of a preliminary project plan, is requisite [20].

Phase 2: Data Understanding
Central to Data Understanding is the acquisition of data. Ideally, an organization already engages in pertinent data collection (e.g., via Internet of Things (IoT) sensor technology) and storage with a database management system. To ascertain if the data is relevant, its information and value needs to be evaluated. The data needs to be described to get a common sense of what it contains and what not. This involves a high-level description that contains certain characteristics of the data such as the volume, its format properties, the number of data points, source and further features. Subsequent analysis possibly unveils first hidden information of the data, such as dominant market segments for specific product groups. In this phase, data quality is evaluated. Criteria for the quality of the data are for example the amount of data or number of missing values. It can also unveil features that are missing in the data at this point. The CRISP-DM model gives an option to circle back to the Business Understanding phase, as understanding the data of a company is closely related to understanding the company itself [20].

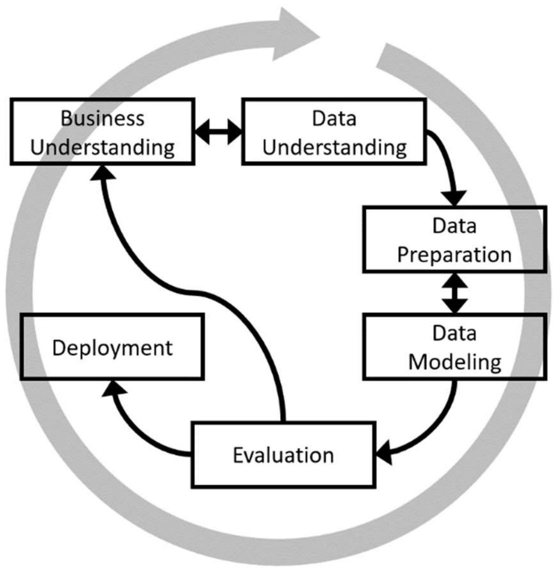

Figure 1: Circular phase model of CRISP-DM [20]

Phase 3: Data Preparation
With this step, the data that is relevant to the data mining objectives must be identified, selected, and subsequently prepared for the modeling phase. Preparation entails data preprocessing, which involves cleaning the data and convert it into a suitable format for modeling. This entails the identification and resolution of redundant attributes, inconsistencies, and outliers. Redundant attributes can be handled by removing duplicated ones, which at the same time achieves data reduction. Reducing the dimension of the data to the information that is relevant without loosing important information is an essential aim of

preprocessing as well. Furthermore, detecting and handling outliers and missing values are crucial for the applicability of the data in Machine Learning models. Another common preprocessing step is to normalize the data to enhance comparability of different features of the data [18], [20].

Phase 4: Modeling
Several models can be considered to solve a problem or execute a certain task. For example, classification tasks might leverage a Decision Tree model or a Neural Network architecture. To find the most suitable model, ideally more than one model is implemented. By benchmarking their performance, the superior model can be identified and deployed. Commonly, model evaluation involves comparing metrics such as error rates or accuracy. The Modeling and Data Preparation phases are interconnected in a cyclic process. Should the model be unable to process data in the preprocessed format, a step backwards to the Data Preparation phase becomes necessary. This iterative process may require multiple cycles to achieve optimal compatibility between the data format and the model's requirements [18], [20].

Phase 5: Evaluation
The assessment of the model must also incorporate an analysis of its contribution to the overarching (business) objectives intended by the initiative. To evaluate the impact of the model on the business objectives, it needs to be implemented and tested. Should the model fail to meet the project objectives, an investigation into the causative factors is needed. During the Evaluation phase, a comprehensive review of the entire process—from the initial phase through to the modeling phase—is conducted to ensure that all critical aspects have been accounted for in the project. If the model does not support the business objectives, the process circles back and starts over in the Business Understanding phase [20].

Phase 6: Deployment
If the model successfully passed the Evaluation phase, it can be deployed to be used for its task in the company. The deployment of an ML model itself is a project that needs to take into account many aspects. Therefore, a deployment plan is needed. Additionally, the model's validity must be regularly assessed. Here, online and offline models can be differentiated. The majority are offline models, trained once using data that is current at that time. Should the data evolve, these models do not automatically adapt unless they are retrained or further trained with new data. Conversely, online models are continuously updated with emerging data, ensuring they remain current and avoid obsolescence [29]. To make the process transparent and understandable for all stakeholders, a documentation is created [20].

2.3. Phase Mapping of VSM and CRISP-DM
From the description of VSM and CRISP-DM that this paper aims to merge, it can already be deducted that they share similarities. Data related technologies support VSM already, but for now there is no structured consideration of how to imply Data Science approaches in VSM, considering it as a holistic Data Mining Project [1], [14], [30]. In the following section, a mapping of the VSM-related phase model and the CRISP-DM phase model is conducted.

The Value Stream Mapping phase encompasses the Business and Data Understanding stages of CRISP-DM, through documenting and analyzing the current Value Stream. This process involves data collection and analysis to refine the Value Stream Map. Value Stream Analysis aligns with CRISP-DM's Data Understanding, Preparation, and Modeling phases, facilitating optimization strategies for Value Stream Design. The Design phase aims to enhance the Value Stream by minimizing waste and reducing lead times, necessitating evaluation to confirm improvements. This is reflected by a part of the Evaluation phase in the CRISP-DM. Subsequently, these optimizations are implemented during the Value Stream Planning phase, paralleling CRISP-DM's Deployment phase, to actualize the refined Value Stream.

2.4. Design of an operational Framework
The design of the operational framework refers to the assignment of VSM phases and CRISP-DM phases, as well as allocating adequate tools of Data Sciences to support the VSM methodology in a data-based manner. The generic structure of the framework is depicted in Figure 2, whereas each procedure including the specific phases represents the horizontal and vertical axes, forming a matrix. The framework aims to emphasize the view on digital VSM as a Data Science initiative. Consequently, this section provides a framework of tools explicitly related to Data Science to cover all phases of VSM.

As the operational framework represents a generic structure, the proposed measures within the matrix represent a selection of best practices, from which the user utilizes the ones, which fulfill the specific requirements and optimally fit to the given data environment.

Value Stream Mapping involves digitally documenting the current value stream, utilizing digital tools such as sensors, ID technologies (e.g., radio-frequency identification (RFID), barcode), and tracking devices as previously mentioned in [1]. The data recording takes place. The creation of a digitized Value Stream Map is facilitated by employing a digital twin or a dedicated dashboard. The second phase in VSM, Value Stream Analysis, aligns with CRISP-DM's Data Understanding, Preparation, and Modeling phases, embodying VSM's most comprehensive task. It contains different steps from the collection of the data to building models. Initially, consolidating data, possibly via database management systems, is crucial for further analysis [31]. To further understand the data, it needs to be analyzed by describing and exploring it. This can be supported by data visualization (e.g. with histograms, scatter- and box plots). Process mining helps as well to consolidate and understand all the process related data. Existing dashboards may aid in data exploration, revealing potential data insufficiencies. In cases of scarce data - complex or costly to gather - special models are employed to extract insights from limited data points [32], [33].

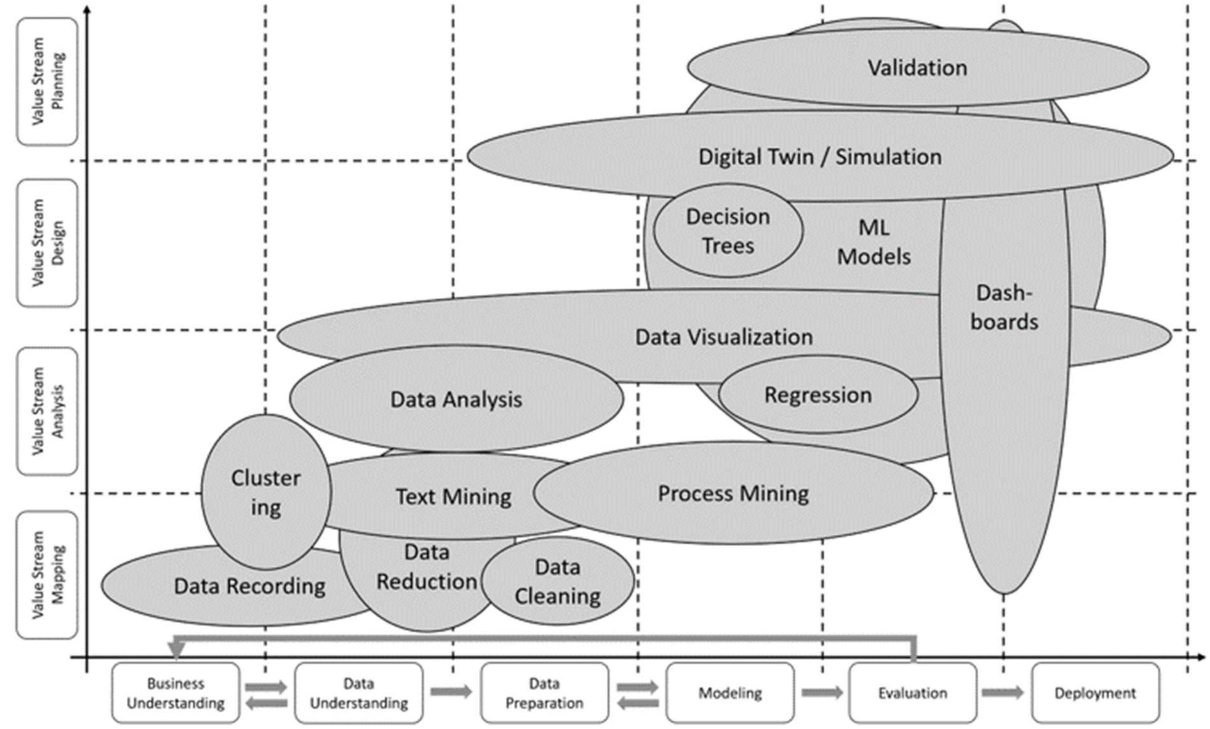

Figure 2: Mapping of the CRISP-DM phases and the VSM phases with exemplary measures

If unstructured data, such as feedback of employees and customers, need to be considered, we need to apply tools to get more structured information from the data, e.g. text mining, natural language processing (NLP) or clustering [34], [35], [36]. These activities occur in the Data Preparation and Understanding phase and extend into the modeling phase as well (e.g. when clustering algorithms are used). The Data Preparation phase covers data cleaning (e.g. noise reduction, handling missing data and outliers), data reduction and data construction (e.g. feature engineering and dimension reduction) [37], [38], [39]. The prepared data is formatted for model testing. Depending on the data and the goal that wants to be achieved, different ML models are of use, e.g. decision trees or a regression. In this Value Stream Analysis phase, ML models help to identify waste causes, predict future bottlenecks, and pinpoint inefficiencies, informing the Value Stream Design phase to craft a more efficient Value Stream.

In the Value Stream Design phase which also intersects with the Modeling phase of the CRISP-DM cycle ML models are used to optimize the Value Stream, e.g. supported by simulations and reinforcement learning with the digital twin [40]. What model can be beneficial to support reducing wastes and therefore optimize the Value Stream is depending on the problem and the data that is available. Therefore, expertise and experienced Data Scientist are needed [19]. The Evaluation phase, involving iterative interactions with the Modeling phase, intersects with both the Value Stream Design and Planning phases. This is reasoned by the iterative implementation of improvement measures of the Value Stream Planning phase as a real life validation of the model. This last phase of the VSM process is mapped with the Deployment phase of the CRISP-DM, where the evaluated model is deployed to be used in VSM.

3. Results and Discussion

This work offers an operational framework that can be used to utilize the full potential of the Value Stream related data and available Data Science applications in the context of VSM. The provided mapping of the phases of CRISP-DM and VSM proves the view on a data-based VSM as a Data Mining project. Data Mining enables the knowledge gaining from available business data in VSM to its full potential. As VSM, CRISP-DM is based on iterations and represents a continuous approach, which leads to a steady improvement of data quality and knowledge gaining with each cycle. To map the phases, the concrete tasks in each phase of the two procedures are reviewed and linked to each other. The operational framework then added tools that can be used in a real world scenario to optimize the data-driven VSM approach.

A main advantage of implementing Data Science applications in VSM is that it enables the possibility for real-time adjustments in the Value Stream. For example, once implemented in a digital twin of the Value Stream which is connected with the real Value Stream via digitization and digitalization technologies (sensors, RFID, barcodes etc.) it can be analyzed and adjusted agile in real-time. Another advantage is that the main KPIs can be observed in real-time as well, offering opportunity for close management. Using ML models that predict potential bottlenecks and identify inefficiencies can further lead to optimized use of resources like workforce, machines and materials. This can also be done by predicting the future demand. Furthermore, Data Science opens opportunities to evaluate the increasing amount of available business data in the area of production and logistics.

The provided framework supports the utilization of data within the VSM methodology. It provides an orientation to merge VSM with the Data Mining process.

4. Limitations and Conclusion

This work offers a holistic view on the VSM and Data Science process. This is resulting from the observed similarities in both procedures and the need of VSM to become more digital and data driven to foster optimizations in a more structured manner compared to the conventional approach. Data Science is a generic term for a lot of models applications and a big research field, a more concrete and process oriented view on it was demanded. The CRISP-DM offers this as a process standard for Data Science projects.

The tools and methodologies pertinent to Data Science disciplines this paper suggest to use extends beyond those exemplarily named within this work. The complexity and effectiveness of these tasks are significantly influenced by the specific use cases to which they are applied, suggesting that a universal approach may not be feasible. Moreover, the application of data-related technologies within a company's infrastructure is primarily contingent upon its existing system landscape. Consequently, the scope of this work does not cover data quality and quantity considerations, though they play a crucial role in the success of Data Mining projects. The process of data collection itself presents another layer of complexity, heavily influenced by the nature of the product and the industry involved. The intricacy of acquiring relevant and high-quality data varies significantly across different sectors, impacting the feasibility and efficiency of data-driven initiatives. Furthermore, it is critical to note that the pursuit of digitization should align with and serve the overarching objectives of an organization, rather than being pursued reasoned by itself. This work adopts a generalist approach and, therefore, does not delve into how VSM and related digital transformation efforts can be tailored to support specific business goals such as enhanced transparency, documentation, and continuous improvement.

In summary, while this study provides valuable insights and suggestions for employing Data Science applications in the domain of VSM, it acknowledges the limitations posed by the complexity of these tasks, the dependency on the organizational context, and the nuances of data quality and acquisition. Future research could benefit from exploring these dimensions in greater detail, focusing on tailored solutions that align technological advances with specific business objectives and industry requirements.

Partially, the considered tools and its utilization are already investigated in terms of VSM by various studies, as mentioned in the previous section. By the provided framework, an overview of Data Science tools to be utilized in the framework VSM is given, based on the view of VSM as a Data Mining process. In the next step, the interplay between the different tools with focus on synergies requires a deeper investigation as well as its validation in operational environments.

5. References

[1] T. Wollert and F. Behrendt, “Future application of VSM in digitalized environments,” 2022, doi: 10.25673/85958.

[2] T. Meudt, J. Metternich, and E. Abele, “Value stream mapping 4.0: Holistic examination of value stream and information logistics in production,” *CIRP Ann.*, vol. 66, no. 1, pp. 413–416, 2017, doi: 10.1016/j.cirp.2017.04.005.

[3] V. Balaji, P. Venkumar, M. S. Sabitha, and D. Amuthaguka, “DVSMS: dynamic value stream mapping solution by applying IIoT,” *Sādhanā*, vol. 45, no. 1, p. 38, Dec. 2020, doi: 10.1007/s12046-019-1251-5.

[4] J. Horsthofer-Rauch, M. Schumann, M. Milde, S. Vernim, and G. Reinhart, “Digitalized value stream mapping: review and outlook,” *Procedia CIRP*, vol. 112, pp. 244–249, 2022, doi: 10.1016/j.procir.2022.09.079.

[5] L. Tomidei, N. Sick, M. Guertler, J. Schallow, D. Lenze, and J. Deuse, “Dynamic Value Stream Mapping—How Industry 4.0 Can Help Us to Learn to See Better,” in *Production Processes and Product Evolution in the Age of Disruption*, F. G. Galizia and M. Bortolini, Eds., in Lecture Notes in Mechanical Engineering. , Cham: Springer International Publishing, 2023, pp. 753–762. doi: 10.1007/978-3-031-34821-1_82.

[6] K. Muehlbauer, M. Wuennenberg, S. Meissner, and J. Fottner, “Data driven logistics-oriented value stream mapping 4.0: A guideline for practitioners,” *IFAC-Pap.*, vol. 55, no. 16, pp. 364–369, 2022, doi: 10.1016/j.ifacol.2022.09.051.

[7] S. Kroeger and M. F. Zaeh, “Towards an Efficient, Comprehensive Value Stream Planning in Production Networks,” *Procedia CIRP*, vol. 107, pp. 782–787, 2022, doi: 10.1016/j.procir.2022.05.062.

[8] T. Teriete, M. Böhm, B. K. Sai, K. Erlach, and T. Bauernhansl, “Event-based Framework for Digitalization of Value Stream Mapping,” *Procedia CIRP*, vol. 107, pp. 481–486, 2022, doi: 10.1016/j.procir.2022.05.012.

[9] B. A. Lameijer, H. Boer, J. Antony, and R. J. M. M. Does, “Continuous improvement implementation models: a reconciliation and holistic metamodel,” *Prod. Plan. Control*, vol. 34, no. 11, pp. 1062–1081, Aug. 2023, doi: 10.1080/09537287.2021.1974114.

[10] J. Mangers, M. Amne Elahi, and P. Plapper, “Digital twin of end-of-life process-chains for a circular economy adapted product design – A case study on PET bottles,” *J. Clean. Prod.*, vol. 382, p. 135287, Jan. 2023, doi: 10.1016/j.jclepro.2022.135287.

[11] A. Lugert, K. Völker, and H. Winkler, “Dynamization of Value Stream Management by technical and managerial approach,” *Procedia CIRP*, vol. 72, pp. 701–706, 2018, doi: 10.1016/j.procir.2018.03.284.

[12] Y. E. Kihel, A. E. Kihel, and S. Embarki, “Optimization of the Sustainable Distribution Supply Chain Using the Lean Value Stream Mapping 4.0 Tool: A Case Study of the Automotive Wiring Industry,” *Processes*, vol. 10, no. 9, p. 1671, Aug. 2022, doi: 10.3390/pr10091671.

[13] M. Bega *et al.*, “Extension of value stream mapping 4.0 for comprehensive identification of data and information flows within the manufacturing domain,” *Prod. Eng.*, vol. 17, no. 6, pp. 915–927, Dec. 2023, doi: 10.1007/s11740-023-01207-5.

[14] R. C. S. Mariappan, A. Veerabathiran, P. K.P, and V. K.E.K, “Intelligent VSM Model: a way to adopt Industry 4.0 technologies in manufacturing industry,” *Int. J. Adv. Manuf. Technol.*, vol. 129, no. 5–6, pp. 2195–2214, Nov. 2023, doi: 10.1007/s00170-023-12406-w.

[15] B. Rahardjo, F.-K. Wang, R.-H. Yeh, and Y.-P. Chen, “Lean Manufacturing in Industry 4.0: A Smart and Sustainable Manufacturing System,” *Machines*, vol. 11, no. 1, p. 72, Jan. 2023, doi: 10.3390/machines11010072.

[16] A. Florescu and S. Barabas, “Development Trends of Production Systems through the Integration of Lean Management and Industry 4.0,” *Appl. Sci.*, vol. 12, no. 10, p. 4885, May 2022, doi: 10.3390/app12104885.

[17] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews,” *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[18] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. in Intelligent Systems Reference Library, vol. 72. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-10247-4.

[19] F. Martinez-Plumed *et al.*, “CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories,” *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021, doi: 10.1109/TKDE.2019.2962680.

[20] R. Wirth and J. Hipp, “CRISP-DM: Towards a Standard Process Model for Data Mining,” Jan. 2000.

[21] K. Erlach, *Value Stream Design: The Way Towards a Lean Factory*. in Lecture Notes in Logistics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. pp. 25-104 doi: 10.1007/978-3-642-12569-0.

[22] M. Thulasi, A. A. Faieza, A. S. Azfanizam, and Z. Leman, “State of the Art of Dynamic Value Stream Mapping in the Manufacturing Industry,” *J. Mod. Manuf. Syst. Technol.*, vol. 6, no. 1, pp. 41–52, Mar. 2022, doi: 10.15282/jmmst.v6i1.7376.

[23] M. Ramadan, M. Alnahhal, and B. Noche, “RFID-Enabled Real-Time Dynamic Operations and Material Flow Control in Lean Manufacturing,” in *Dynamics in Logistics*, H. Kotzab, J. Pannek, and K.-D. Thoben, Eds., in Lecture Notes in Logistics. , Cham: Springer International Publishing, 2016, pp. 281–290. doi: 10.1007/978-3-319-23512-7_27.

[24] H. Wang, Q. He, Z. Zhang, T. Peng, and R. Tang, “Framework of automated value stream mapping for lean production under the Industry 4.0 paradigm,” *J. Zhejiang Univ.-Sci. A*, vol. 22, no. 5, pp. 382–395, May 2021, doi: 10.1631/jzus.A2000480.

[25] C. Oberhausen, M. Minoufekr, and P. Plapper, “Continuous improvement of complex process flows by means of stream as the ‘standardized cross-enterprise value stream management method,’” in *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore: IEEE, Dec. 2017, pp. 460–464. doi: 10.1109/IEEM.2017.8289933.

[26] I. Göpfert, D. Braun, and M. Schulz, Eds., *Automobillogistik*. Wiesbaden: Springer Fachmedien Wiesbaden, 2017. pp. 273-274 doi: 10.1007/978-3-658-11103-8.

[27] “DIN ISO 22468:2022-07, Wertstrommethode_(VSM) (ISO_22468:2020).” Beuth Verlag GmbH. doi: 10.31030/3331449.

[28] DIN and DKE, “Deutsche Normungsroadmap Künstliche Intelligenz (Ausgabe 2),” Berlin, Jan. 2022. [Online]. Available: www.din.de/go/normungsroadmapki

[29] S. C. H. Hoi, D. Sahoo, J. Lu, and P. Zhao, “Online learning: A comprehensive survey,” *Neurocomputing*, vol. 459, pp. 249–289, Oct. 2021, doi: 10.1016/j.neucom.2021.04.112.

[30] T. Wollert and F. Behrendt, “Application potentials of a data-based Value Stream Map model in production and logistics,” 2023, doi: 10.25673/103493.

[31] M. L. Gillenson, *Fundamentals of database management systems*. John Wiley & Sons, 2023.

[32] C.-A. Holst and V. Lohweg, “Scarce Data in Intelligent Technical Systems: Causes, Characteristics, and Implications,” *Sci*, vol. 4, no. 4, p. 49, Dec. 2022, doi: 10.3390/sci4040049.

[33] J. Kösters, M. Schöne, and M. Kohlhase, “Benchmarking of Machine Learning Models for Tabular Scarce Data,” Center for Applied Data Science, Bielefeld University of Applied Science, Gütersloh, Germany, 3729, 2023. [Online]. Available: https://www.hsbi.de/publikationsserver/record/3729

[34] S. Jain, A. De Buitleir, and E. Fallon, “A Review of Unstructured Data Analysis and Parsing Methods,” in *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India: IEEE, Mar. 2020, pp. 164–169. doi: 10.1109/ESCI48226.2020.9167588.

[35] M. Y. Saeed, M. Awais, R. Talib, and M. Younas, “Unstructured Text Documents Summarization With Multi-Stage Clustering,” *IEEE Access*, vol. 8, pp. 212838–212854, 2020, doi: 10.1109/ACCESS.2020.3040506.

[36] A. Gupta, R. Singh, V. K. Nassa, R. Bansal, P. Sharma, and K. Koti, “Investigating Application and Challenges of Big Data Analytics with Clustering,” in *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, Coimbatore, India: IEEE, Oct. 2021, pp. 1–6. doi: 10.1109/ICAECA52838.2021.9675483.

[37] C. R. Turner, A. Fuggetta, L. Lavazza, and A. L. Wolf, “A conceptual basis for feature engineering,” *J. Syst. Softw.*, vol. 49, no. 1, pp. 3–15, 1999.

[38] P. Cunningham, “Dimension reduction,” in *Machine learning techniques for multimedia: Case studies on organization and retrieval*, Springer, 2008, pp. 91–112.

[39] A. Maćkiewicz and W. Ratajczak, “Principal components analysis (PCA),” *Comput. Geosci.*, vol. 19, no. 3, pp. 303–342, 1993.

[40] K. Xia *et al.*, “A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence,” *J. Manuf. Syst.*, vol. 58, pp. 210–230, Jan. 2021, doi: 10.1016/j.jmsy.2020.06.012.