Enterprise-Architecture-Data-Science Modeling Framework for Data Asset Valuation

Matthias Pohl[1,2]

[1]German Aerospace Center (DLR), Institute of Data Science, Mälzerstr. 3-5, 07745 Jena, Germany
[2]Otto von Guericke University, Faculty of Computer Science, Universitatsplatz 2, 39106 Magdeburg, Germany

Abstract

Organizations struggle to realize value from artificial intelligence investments, with 74-76% of companies failing to achieve significant returns. This research addresses two critical questions: how to integrate data science initiatives into enterprise architecture (EA) modeling to connect value realization, and how to assess data asset value in a structured manner. The paper discusses a comprehensive framework that combines EA principles with data science methodologies, incorporating TOGAF's architectural layers with performance indicator ontologies. Using predictive maintenance in smart manufacturing as a demonstration case, the framework links Key Performance Indicators (KPIs) with data science model evaluation metrics through contingency table analysis. The approach enables organizations to quantify data asset value by comparing model performance against baseline scenarios, translating technical metrics like accuracy into business-relevant indicators such as maintenance costs. This integration provides a systematic methodology for valuing data assets and demonstrating the business impact of data science initiatives.

Keywords

Enterprise Architecture Modeling, Data Valuation, Smart Manufacturing, Key Performance Indicator, Ontology

1. Introduction

Research in the consulting business shows that 74% to 76% of companies do not get real value from their investments in artificial intelligence (AI). Only 4% to 5% of these companies find significant value from their AI efforts on a large scale [1]. On the other hand, companies that use AI successfully report revenue growth that is 2 to 5 times higher than those that do not utilize AI effectively. They also achieve profitability that is 40% to 60% greater than their peers. Additionally, they can see EBITDA gains of up to 25% [1, 2]. Further, many companies that are slow to adopt AI are facing significant financial challenges. Approximately 99% of these organizations are experiencing losses due to AI-related risks, with an average loss of around $4.4 million each [3].

A report from McKinsey reveals that 78-79% of companies utilizing generative AI do not experience a significant impact on their profits. This situation leads to what is referred to as the "gen AI paradox," where extensive adoption of this technology yields minimal returns. Only 8% of organizations are effectively scaling their AI initiatives, and fewer than 10% of AI projects advance beyond the pilot phase [4, 5]. Moreover, despite a sixfold increase in enterprise AI spending, only 27% of organizations have fully integrated AI into their operations. This indicates that 73% of companies have struggled to translate their investments into practical AI applications [6].

McKinsey indicates that companies waste 70% of their efforts on data cleansing, with more than half of data lakes not being suitable for their intended purposes [7]. Boston Consulting Group (BCG) highlights that only 38% of organizations have established a strong data-driven culture, and 74% face ongoing challenges in effectively integrating big data into their operations. Most executives recognize this issue, with 91% citing difficulties related to people and processes as their primary barriers, rather than technology [8]. McKinsey also points out that the lack of a clear AI strategy is the main barrier to

PoEM2025: Companion Proceedings of the 18th IFIP Working Conference on the Practice of Enterprise Modeling: PoEM Forum, Doctoral Consortium, Business Case and Tool Forum, Workshops, December 3-5, 2025, Geneva, Switzerland
matthias.pohl@dlr.de (M. Pohl)
0000-0002-6241-7675 (M. Pohl)

adoption. Only 17% of companies have successfully identified and mapped potential AI opportunities within their organizations [9].

Former research reveals that an organization's ability to adopt AI is influenced by its internal capabilities, which are shaped by its enterprise architecture (EA). To achieve successful AI integration, companies should ensure a strong alignment between their AI initiatives and business units. This alignment is crucial to ensure that AI strategies support business goals and objectives while also considering external factors such as market trends and competition [10].

Recent research from leading consulting firms emphasized that data is a crucial element in the effective deployment of AI. AI is primarily associated with the use of large, pre-trained neural networks designed for specific use cases. These advanced models are often part of extensive data science projects, utilizing deep learning techniques to extract valuable insights and predictions from substantial datasets. Furthermore, terms such as data science, data analytics, data mining, and machine learning are commonly used interchangeably in the context of AI. Each of these domains plays an essential role in deriving knowledge from data [11, 12]. Collectively, these fields contribute significantly to the progress and application of AI technologies across diverse industries.

Valuing data is crucial for understanding and utilizing it effectively as an asset. This process aids companies in making informed decisions regarding investments in technology, marketing, and research [13]. As organizations collect increasing amounts of data, it is vital to comprehend its economic value to guide their choices. By assessing data, businesses can allocate resources strategically and invest in areas that yield the best returns. Data is increasingly recognized as an asset that can appear on a company's balance sheet. An accurate valuation of data can enhance the overall worth of the company. Moreover, a clear understanding of data value enables companies to identify and mitigate risks associated with data usage. This insight helps improve risk assessment and informs strategies for managing potential impacts on the business [14, 15].

In this short paper, after introducing the research approach, we will first explain how to integrate data science projects into enterprise architecture (EA) modeling. Next, we will introduce the concept of ontology, which is important for managing key performance indicators (KPIs). We will show how organizations use key performance indicators together with data science model evaluation to evaluate and assess the value of data assets.

2. Research Approach

As EA plays a pivotal role, the first aim of this research is to examine in the integration of data science initiatives into EA. It focuses on developing a comprehensive modeling approach that is closely aligned with EA principles. In this framework, data is regarded not only as a fundamental asset but also as a vital driver of value for organizations. The design and structure of statistical and machine learning models are significantly shaped by the quality and characteristics of the data involved. As a result, the success and value derived from data science initiatives are inherently tied to the quality and assessment of the data used. To explore this essential relationship, the research will investigate various methodologies for valuing data assets within the context of data science.

The following research questions are addressed:

- **RQ1:** How can data science initiatives be integrated into EA modeling to connect the value realization from these initiatives in organizations?
- **RQ2:** How can the value of data assets related to data science initiatives be assessed in a structured manner?

2.1. Methodology

We are employing a design science research methodology, which prioritizes the systematic development and evaluation of innovative artifacts to address intricate challenges within the domain [16, 17, 18]. At this juncture, we have delineated the objectives of our proposed solution, articulating specific aims that we seek to accomplish through this research endeavor. Additionally, we have constructed an initial conceptual framework that encapsulates our approach and serves as a foundational basis for further investigation and iterative refinement of the solution (see Section 3).

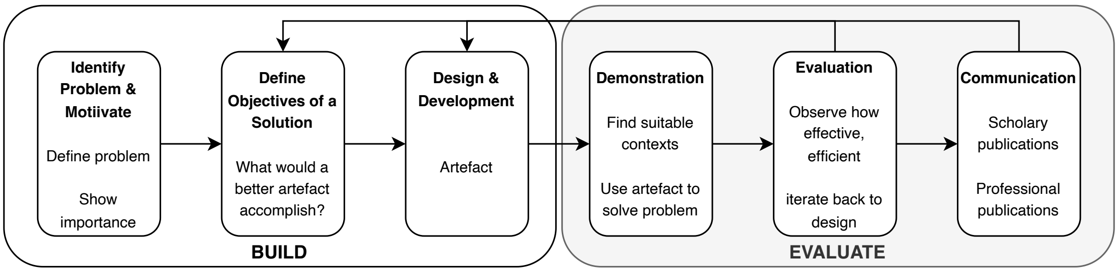

Figure 1: Design Science Research Approach according to [18].

3. Current Research Work

Data science can be viewed as a systematic and structured approach aimed at extracting meaningful information from various datasets, ultimately facilitating the discovery of valuable insights and knowledge [19, 11]. Within this field, terms such as data analytics, data mining, machine learning, and artificial intelligence frequently emerge. While these terms can sometimes be used interchangeably, they also represent distinct components of the broader discipline of data science.

The applications of data science can be categorized into three main types: descriptive, predictive, and prescriptive [20]. Descriptive applications primarily focus on uncovering trends and patterns within data. This includes methodologies like clustering, which groups similar data points, as well as pattern recognition projects that seek to identify significant relationships or anomalies within datasets. Predictive applications leverage statistical techniques and algorithms to forecast future outcomes based on historical data. This category can be further subdivided into classification approaches, which assign data to predefined categories, and regression approaches, which model the relationships between variables to predict continuous outcomes. Finally, prescriptive applications utilize advanced methodologies, such as simulation studies and optimization techniques. They provide actionable recommendations and strategies to guide decision-making, enabling organizations to achieve their goals effectively and efficiently.

The research at hand focuses on the dynamics of decision-making in business processes, emphasizing the role of classification and optimization methodologies.

3.1. Holistic View on Integrated Data Science Processes in EA

In our initial research, we conducted a thorough literature review on the integration of data science projects within EA. The analysis of existing studies on incorporating DS processes into EA reveals a notable deficiency in providing a comprehensive perspective on the implications of DA. Currently available methodologies tend to focus either on the business architecture layer [21, 22, 23, 24, 25, 26], which includes conceptual data science processes, or adopt a high-level viewpoint that fails to account for the specific details of DS processes and infrastructure [27, 28, 29].

In light of the findings, we propose a combined approach. This framework seeks to integrate the various layers of architecture as delineated in established EA frameworks, such as The Open Group Architecture Framework (TOGAF) [30], with pertinent data science methodologies [19]. The objective is to develop a cohesive understanding of how DS can be effectively incorporated into the overall architecture, thereby ensuring alignment between both strategic and operational elements (see Figure 2).

The comprehensive architectural approach can be articulated with precision as follows: Within the framework of TOGAF, enterprise design is organized into three primary layers. The Business Architecture layer delineates the fundamental business processes, while the Technology Architecture encompasses the hardware and software technologies essential for supporting enterprise operations, thereby providing the requisite infrastructure for effective business and systems integration. Importantly, the Information Systems layer is further subdivided into two critical components: Data Architecture and Application Architecture. Data Architecture centers on the systematic organization and management of data assets to facilitate business processes and informed decision-making. Conversely, Application Architecture pertains to the design and interrelation of software applications within the enterprise, with a focus on their alignment with organizational objectives and their seamless integration.

The Business Architecture is presented through a detailed diagram that consists of two main components. The left side illustrates a typical structure of business processes, including an event-driven diagram and a business unit diagram [30]. This section highlights a central business process essential for enterprise operations, supported by functions and services that facilitate effective execution and alignment with the organization's objectives. The business process serves as a mechanism for transforming an input, typically a product or service, into a valuable output. This transformation is initiated by an actor responding to a specific business event, which triggers the process workflow. Each business process is managed by an internal actor usually associated with a specific business division. On the right side of the architecture, a foundational model of a DS service is depicted, overseen by a Data Scientist within either the IT or DS division. This service demonstrates how DS can enhance decision-making and operational efficiency across the organization. The success of a DS service often relies on a framework of multiple DS business processes. These processes are guided by established DS methodologies, which provide a systematic approach to data analysis and interpretation. By following these methodologies, organizations can ensure that their DS initiatives are effective and aligned with broader business goals, ultimately leading to valuable insights and innovative solutions (see [21, 22, 23, 24, 25, 26]).

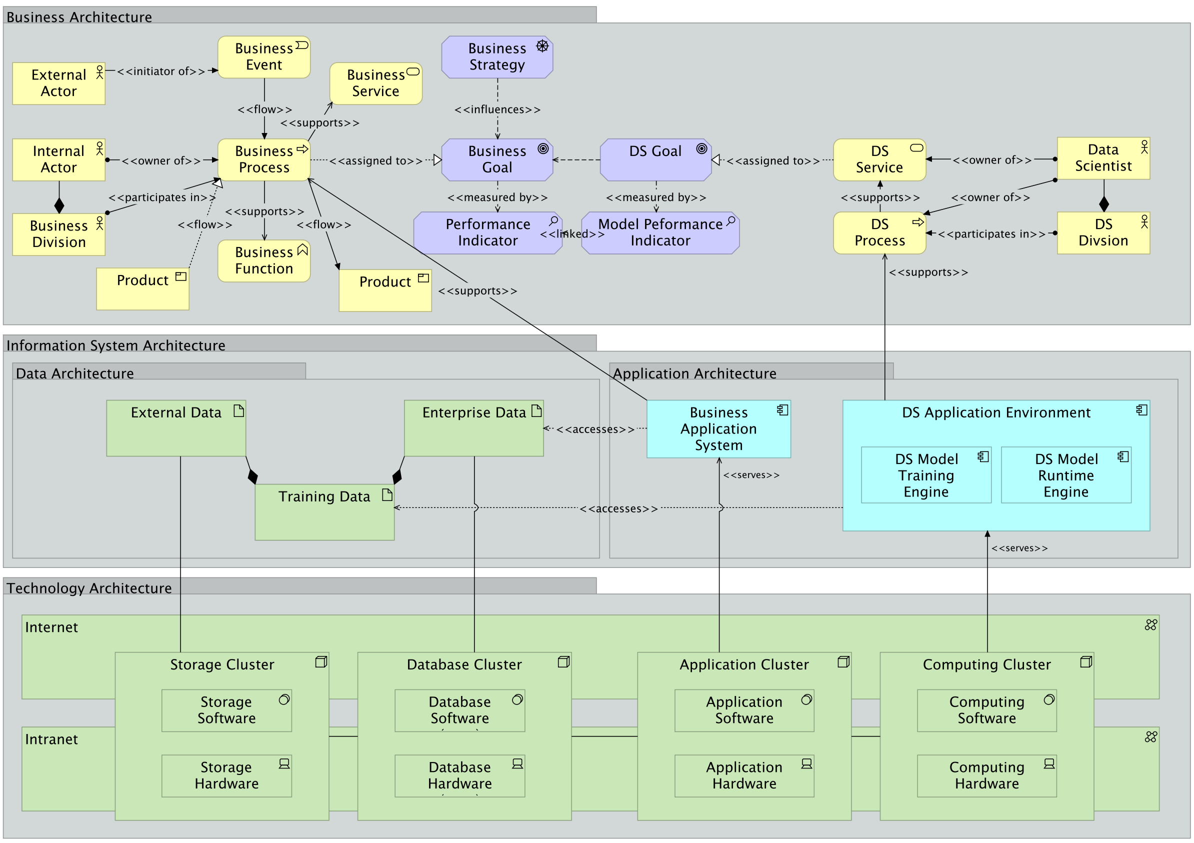

Figure 2: Holistic View on Integrated DS Processes in EA.

3.2. Ontology of Performance Indicators

In the realm of performance measurement, various concepts are employed to model the intricate interactions among Performance Indicators [31, 32]. A notable model within this field is KPIOnto, which articulates indicators as mathematical expressions that integrate input parameters and an aggregation function [33, 34, 35]. These indicators are intentionally aligned with specific business objectives to ensure they contribute to the overarching success of the organization. Furthermore, they are interconnected with various dimensions, including business processes, product characteristics, and temporal elements (see Figure 3). Although there have been initiatives to implement ontological frameworks for performance indicators within the manufacturing sector, such efforts remain limited in their widespread acceptance [36]. The potential advantages of employing such ontologies include the standardization of terminology and the enhancement of communication regarding performance metrics. However, further validation is imperative to evaluate their effectiveness comprehensively.

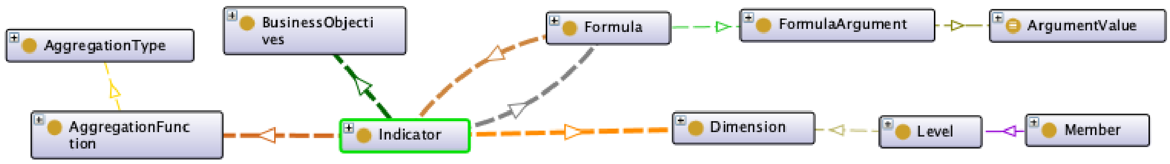

Figure 3: Excerpt of KPI Ontology according to [32].

3.3. Conceptual Valuation Model

According to Decision Theory [37], we can effectively map how data science applications integrate into the decision-making processes within business environments. By examining a range of possible actions, denoted as $a \in A$, alongside various decision states represented by $x \in X$, we can clearly define a specific decision problem, referred to as $D =< A, X >$

These concepts are further reinforced by an information structure, denoted as $I = \{Y, p(x, y)\}$. The information structure encapsulates the outcomes generated by data science applications ($y \in Y$) and a probabilistic relationship between the states and the data ($p(x, y)$). When we consider classification applications as fundamental tools for decision support, we recognize a direct equivalence to the decision problem outlined in Decision Theory.

The anticipated payoff associated with a specific decision-making problem in the absence of prior knowledge is defined as the expected value of a designated measure function (π).

Within the use of data science applications, the optimal outcome is characterized by maximizing the expected value across all conceivable sets of actions. Moreover, when assessing the collective influence of posterior information obtained from data science applications, one can ascertain its value by examining the difference in expected payoffs.

$$V(I) = E_Y E_{Y|X}(\pi(x, a_y)) - E_X(\pi(x, a)) \tag{1}$$

The theoretical implications of this consideration prompt an examination of the complexities involved in measuring decision-making processes. In this context, performance indicators play a pivotal role, acting as economic benchmarks that quantify actions in relation to specific decision problems. By assessing these measured actions against the results obtained from data science applications, we can determine the true value of the information. This is achieved by comparing actions taken under informed conditions with those made without prior knowledge, thereby illustrating the significant influence that data-driven insights can exert on decision outcomes.

4. Discussion & Challenges

It is essential to recognize that the overall value derived from data assets can be significantly enhanced by integrating multiple data science use cases. Hence, it is imperative to expand the framework to accommodate additional use case integration.

To support this expansion, we have developed a use case catalog that serves as a foundational resource, establishing the groundwork for exploring and incorporating new use cases that more effectively leverage existing data assets [38].

However, the intersection of Enterprise Architecture (EA) modeling and KPI ontology introduces considerable complexity. This situation raises essential questions about whether a straightforward mapping or a referencing approach will suffice in creating connections between these two domains. Moreover, we need to assess whether an evaluation of the EA modeling is necessary, leading to inquiries regarding whether this should be treated as an artifact. The question of whether a meta-model is required also arises as a key point of exploration.

Additionally, the valuation step in this framework becomes increasingly complex, as it must address various data science use cases and models. The reliance on a well-defined ontology becomes crucial in navigating this complexity, as it fundamentally impacts the valuation process. Consequently, the complexities involved in evaluating these diverse data science applications are currently under thorough investigation to ensure a robust and effective valuation methodology.

Declaration on Generative AI

During the preparation of this work, the author used Grammarly in order to: Grammar and spelling check, Paraphrase, and reword. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] Boston Consulting Group, The Widening AI Value Gap, Technical Report, Boston Consulting Group, 2025. URL: https://media-publications.bcg.com/The-Widening-AI-Value-Gap-Sept-2025.pdf, accessed: 2025-10-12.

[2] Boston Consulting Group, Where's the value in ai?, 2024. URL: https://www.bcg.com/publications/2024/wheres-value-in-ai, accessed: 2025-10-12.

[3] EY, Ey survey: Companies advancing responsible ai governance linked to better business outcomes, Press Release, 2025. URL: https://www.ey.com/en_gl/newsroom/2025/10/ey-survey-companies-advancing-responsible-ai-governance-linked-to-better-business-outcomes, accessed: 2025-10-12.

[4] McKinsey & Company, The State of AI: How Organizations Are Rewiring to Capture Value, Technical Report, McKinsey & Company, 2025. URL: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai, accessed: 2025-10-12.

[5] McKinsey & Company, Seizing the agentic ai advantage, 2025. URL: https://www.mckinsey.com/capabilities/quantumblack/our-insights/seizing-the-agentic-ai-advantage, accessed: 2025-10-12.

[6] Roland Berger, Navigating the Path to AI-Driven Value Creation, Technical Report, Roland Berger, 2024. URL: https://content.rolandberger.com/hubfs/07_presse/Roland%20Berger_Study_Gen%20AI_x_Data%20Management_final.pdf, accessed: 2025-10-12.

[7] McKinsey & Company, Ten red flags signaling your analytics program will fail, 2024. URL: https://www.mckinsey.com/capabilities/quantumblack/our-insights/ten-red-flags-signaling-your-analytics-program-will-fail, accessed: 2025-10-12.

[8] KPMG, Kpmg ai quarterly pulse survey, 2025. URL: https://kpmg.com/us/en/articles/2025/ai-quarterly-pulse-survey.html, accessed: 2025-10-12.

[9] McKinsey & Company, Ai adoption advances, but foundational barriers remain, 2024. URL: https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain, accessed: 2025-10-12.
[10] P. Stecher, M. Pohl, K. Turowski, Enterprise architecture's effects on organizations' ability to adopt artificial intelligence - A resource-based perspective, in: ECIS 2020 Proceedings, 2020.
[11] M. Pohl, C. Haertel, D. Staegemann, K. Turowski, Data Science Methodology:, in: J. Wang (Ed.), Encyclopedia of Data Science and Machine Learning, IGI Global, 2022, pp. 1201–1214. URL: https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-7998-9220-5.ch070. doi:10.4018/978-1-7998-9220-5.ch070.
[12] C. Haertel, M. Pohl, A. Nahhas, D. Staegemann, K. Turowski, Toward a Lifecycle for Data Science: A Literature Review of Data Science Process Models, in: PACIS 2022 Proceedings, AIS, 2022.
[13] P. Parvinen, E. Pöyry, R. Gustafsson, M. Laitila, M. Rossi, Advancing data monetization and the creation of data-based business models, Communications of the association for information systems 47 (2020) 25–49.
[14] V. Mayer-Schönberger, K. Cukier, Big data: A revolution that will transform how we live, work, and think, Houghton Mifflin Harcourt, 2013.
[15] D. Laney, Infonomics: How to Monetize - Manage and Measure İnformation as an Asset for Competitive Advantage, Routlegde, New York, 2018.
[16] A. R. Hevner, S. T. March, J. Park, S. Ram, Design Science in Information Systems Research, MIS Quarterly 28 (2004) 75–105.
[17] A. R. Hevner, S. Chatterjee, Design research in information systems: theory and practice, number v. 22 in Integrated series in information systems, Springer, New York ; London, 2010. OCLC: ocn471801169.
[18] K. Peffers, T. Tuunanen, M. A. Rothenberger, S. Chatterjee, A design science research methodology for information systems research, Journal of management information systems 24 (2007) 45–77.
[19] C. Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Journal of data warehousing 5 (2000) 13–22.
[20] S. Nalchigar, E. Yu, Business-driven data analytics: A conceptual modeling framework, Data & Knowledge Engineering 117 (2018) 359–372.
[21] H. Takeuchi, K. Imazaki, N. Kuno, T. Doi, Y. Motohashi, Constructing reusable knowledge for machine learning projects based on project practices, Intelligent Decision Technologies 16 (2022) 725–735. URL: https://journals.sagepub.com/doi/full/10.3233/IDT-220252. doi:10.3233/IDT-220252.
[22] H. Takeuchi, J. H. Husen, H. T. Tun, H. Washizaki, N. Yoshioka, Enterprise Architecture-based Metamodel for a Holistic Business—IT Alignment View on Machine Learning Projects, in: 2023 IEEE International Conference on e-Business Engineering (ICEBE), IEEE, Sydney, Australia, 2023, pp. 8–15. doi:10.1109/ICEBE59045.2023.00013.
[23] H. Takeuchi, T. Doi, H. Washizaki, S. Okuda, N. Yoshioka, Enterprise Architecture based Representation of Architecture and Design Patterns for Machine Learning Systems, in: 2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW), IEEE, Gold Coast, Australia, 2021, pp. 245–250. doi:10.1109/EDOCW52865.2021.00055.
[24] H. Takeuchi, J. H. Husen, H. T. Tun, H. Washizaki, N. Yoshioka, Enterprise architecture-based metamodel for machine learning projects and its management, Future Generation Computer Systems 161 (2024) 135–145. doi:10.1016/j.future.2024.06.062.
[25] H. Takeuchi, H. Kaiya, H. Nakagawa, S. Ogata, Practice-based Collection of Bad Smells in Machine Learning Projects, Procedia Computer Science 225 (2023) 517–526. doi:10.1016/j.procs.2023.10.036.
[26] H. Takeuchi, A. Ichitsuka, T. Iino, S. Ishikawa, K. Saito, Assessment Method for Identifying Business Activities to be Replaced by AI Technologies, Procedia Computer Science 192 (2021) 1601–1610. doi:10.1016/j.procs.2021.08.164.
[27] S. Idowu, D. Struber, T. Berger, EMMM: A Unified Meta-Model for Tracking Machine Learning Experiments, in: 2022 48th Euromicro Conference on Software Engineering and Advanced

Applications (SEAA), IEEE, Gran Canaria, Spain, 2022, pp. 48–55. doi:10.1109/SEAA56994.2022.00016.
[28] B. Geerdink, A reference architecture for big data solutions - introducing a model to perform predictive analytics using big data technology, International Journal of Big Data Intelligence 2 (2015) 236. URL: http://www.inderscience.com/link.php?id=72160. doi:10.1504/IJBDI.2015.072160.
[29] J. Husen, H. Washizaki, N. Yoshioka, H. Tun, Y. Fukazawa, H. Takeuchi, Metamodel-Based Multi-View Modeling Framework for Machine Learning Systems:, in: Proceedings of the 11th International Conference on Model-Based Software and Systems Engineering, SCITEPRESS - Science and Technology Publications, Lisbon, Portugal, 2023, pp. 194–201. doi:10.5220/0011699600003402.
[30] P. Desfray, G. Raymond, Modeling enterprise architecture with TOGAF: a practical guide using UML and BPMN, Morgan Kaufmann, Amsterdam ; Boston, 2014.
[31] C. Diamantini, T. Khan, D. Potena, E. Storti, Semantic Models of Performance Indicators: A Systematic Survey, ACM Computing Surveys 57 (2025) 1–37. doi:10.1145/3719291.
[32] M. Del Mar Roldán-García, J. García-Nieto, A. Maté, J. Trujillo, J. F. Aldana-Montes, Ontology-driven approach for KPI meta-modelling, selection and reasoning, International Journal of Information Management 58 (2021). doi:10.1016/j.ijinfomgt.2019.10.003.
[33] C. Diamantini, D. Potena, E. Storti, H. Zhang, An Ontology-Based Data Exploration Tool for Key Performance Indicators, in: R. Meersman, H. Panetto, T. Dillon, M. Missikoff, L. Liu, O. Pastor, A. Cuzzocrea, T. Sellis (Eds.), On the Move to Meaningful Internet Systems: OTM 2014 Conferences, volume 8841, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 727–744. doi:10.1007/978-3-662-45563-0_45, series Title: Lecture Notes in Computer Science.
[34] C. Diamantini, D. Potena, E. Storti, SemPI: A semantic framework for the collaborative construction and maintenance of a shared dictionary of performance indicators, Future Generation Computer Systems 54 (2016) 352–365. doi:10.1016/j.future.2015.04.011.
[35] C. Diamantini, D. Potena, E. Storti, Extended drill-down operator: Digging into the structure of performance indicators, Concurrency and Computation: Practice and Experience 28 (2016) 3948–3968. doi:10.1002/cpe.3726.
[36] G. Pintzos, M. Matsas, G. Chryssolouris, Defining Manufacturing Performance Indicators Using Semantic Ontology Representation, Procedia CIRP 3 (2012) 8–13. doi:10.1016/j.procir.2012.07.003.
[37] D. B. Lawrence, The economic value of information, Springer Science & Business Media, 2012.
[38] M. Pohl, C. Haertel, D. Staegemann, K. Turowski, Categorization of data analytics projects in smart manufacturing, in: 2024 IEEE International Conference on Big Data (BigData), 2024, pp. 3153–3157. doi:10.1109/BigData62323.2024.10825954.