Bus Inf Syst Eng
https://doi.org/10.1007/s12599-025-00940-8

RESEARCH PAPER

Selecting Data Assets in Data Marketplaces

Leveraging Machine Learning and Explainable AI for Value Quantification

Dominik Martin · Daniel Heinz · Moritz Glauner · Niklas Kühl

Received: 23 November 2023 / Accepted: 4 February 2025

Abstract In the era of digital transformation, data is a critical asset driving innovation and competitive advantage for businesses. Data marketplaces have emerged as a key solution for data sharing, yet they face significant challenges, including competitive concerns, matchmaking between data providers and consumers, and a lack of appropriate market mechanisms. This study introduces a data asset value quantification and selection mechanism (DQSM) as an innovative feature for data marketplaces to address these concerns. The DQSM uses Machine Learning and Explainable AI methods to assess the value of data assets, aiding consumers in making informed purchasing decisions. This mechanism addresses the inherent complexities of data asset valuation and selection, thereby increasing marketplace efficiency. Using a design science research approach, the study identifies design principles for the development of the DQSM as a feature of data marketplaces, which are validated through technical experiments with industry and public datasets, as well as interviews with experts in this field. The findings highlight the potential of the DQSM to optimize the discovery and implementation of viable data sharing use cases and to incentivize the adoption of data marketplaces, thereby contributing to more viable and sustainable data ecosystems.

Keywords Data sharing · Data marketplaces · Digital platforms · Explainable artificial intelligence · Feature selection · Design science research

1 Introduction

The digital transformation of industries has positioned data as a fundamental asset for driving innovation, enhancing operational efficiency, and securing competitive advantage (Wixom and Ross 2017). Organizations increasingly recognize not just the value of accumulating data but the strategic imperative of monetizing it (Spiekermann 2019). This shift has led to the emergence of data marketplaces (Fassnacht et al. 2024) – platforms like Dawex,[1] Snowflake Marketplace,[2] and Caruso[3] – designed to facilitate the buying, selling, and sharing of data assets across various industries (Koutroumpis et al. 2017; Spiekermann 2019). These marketplaces aim to streamline the discovery and acquisition of diverse datasets, fostering data-driven innovation and cross-organizational collaboration.

Despite their potential, data marketplaces face a significant hurdle: information asymmetry between data providers and consumers (Zhang et al. 2023; Huang et al. 2021). Unlike physical goods, where quality can often be assessed prior to purchase, data assets are intangible, and their value is highly dependent on the specific context in

[1] https://www.dawex.com/.

[2] https://www.snowflake.com/en/data-cloud/marketplace/.

[3] https://caruso-dataplace.com/.

Accepted after 2 revisions by Natalia Kliewer.

D. Martin · D. Heinz (✉) · M. Glauner
Karlsruhe Institute of Technology, Kaiserstr. 89,
76133 Karlsruhe, Germany
e-mail: daniel.heinz@kit.edu

D. Martin
Trelleborg Sealing Solutions, Schockenriedstr. 1,
70565 Stuttgart, Germany

N. Kühl
University of Bayreuth, Wittelsbacherring 10, 95444 Bayreuth,
Germany

which they are used (Spiekermann 2019). Data consumers struggle to evaluate the quality and relevance of datasets without first accessing them, while data providers are hesitant to reveal detailed information due to concerns over data sovereignty, intellectual property rights, and competitive disadvantage (Enders et al. 2020). This dilemma reflects the traditional ‘Arrow Information Paradox’ (Arrow 1962), where the value of information cannot be fully assessed until after it has been disclosed, at which point the seller has nothing left to sell.

This information asymmetry leads to market inefficiencies in data marketplaces. Data consumers may be reluctant to purchase datasets due to uncertainty about their utility, resulting in decreased transaction volumes and underutilization of potentially valuable data assets (Spiekermann 2019; Sterk et al. 2023). Traditional market mechanisms, such as upfront pricing and detailed product descriptions, are insufficient to bridge this gap, as they either fail to provide the necessary insights or require data providers to disclose sensitive information (Badewitz et al. 2022). For example, a company aiming to enhance its predictive analytics models might consider acquiring external datasets from a data marketplace like Snowflake. Without the ability to assess how these datasets will integrate with their existing models or improve predictive accuracy, the company faces significant risk. Purchasing datasets that ultimately do not contribute value leads to wasted resources and erodes trust in the marketplace platform.

Addressing this challenge requires innovative solutions that enable data consumers to evaluate the potential value of datasets prior to purchase without compelling data providers to expose proprietary information (Fassnacht et al. 2024). One promising approach lies in the application of methods from machine learning (ML) and explainable artificial intelligence (XAI). Both methods have the potential for abstract representations and scores that indicate their potential utility for specific use cases, all while preserving the confidentiality of raw data (Sahakyan et al. 2021; Hirt et al. 2023).

In this context, we introduce the concept of a *data asset value quantification and selection mechanism (DQSM)*. The DQSM is envisioned as a technical feature within data marketplaces that provides an assessment of data assets’ value relative to a consumer’s specific analytical needs. By employing ML and XAI techniques, the DQSM can offer insights into how a dataset might contribute to a particular model or analysis without revealing the underlying data. This mechanism aims to empower data consumers with better decision-making tools, reduce uncertainty, and enhance trust in data marketplaces, ultimately also helping data providers and data marketplaces through increased usage frequency and trust.

Consider the scenario of an insurance company developing a pay-how-you-drive pricing model that uses an analytical model to assess claims risk based on a driver’s real-time behavior. Traditionally, this model relies on collecting detailed driving behavior data through retrofitted sensors, which is costly and can raise privacy concerns. Alternatively, the company could purchase existing datasets such as in-car telemetry (e.g., acceleration, GPS locations, fuel consumption), weather conditions, and traffic data from a data marketplace to build a virtual driving behavior sensor. The DQSM would allow the company to evaluate which datasets are most likely to improve its risk assessment models, facilitating a targeted and efficient data acquisition strategy without unnecessary expense or data exposure.

The objective of this study is to develop and validate the DQSM as a solution to the problem of information asymmetry in data marketplaces – enabling the quantification of the value of data assets for informed consumer purchasing decisions without the need for data providers to reveal raw data to consumers. Guided by the design science research (DSR) paradigm (March and Smith 1995; Hevner et al. 2004; Venable et al. 2016), we seek to create an artifact that not only addresses the practical challenges faced by data consumers and providers but also contributes to the theoretical understanding of data exchange mechanisms. In doing so, we derive design knowledge represented by design principles (DPs) to support the design of a DQSM as a technical feature of a data marketplace.

This leads us to the primary research question: *What design knowledge guides the development of a DQSM to alleviate information asymmetry in data marketplaces?* To systematically explore this question, we break it down into the following sub-questions:

RQ1: How can ML and XAI methods be leveraged to assess the potential value of data assets for specific use cases within a data marketplace?

RQ2: What factors should be considered in designing an effective data acquisition strategy supported by a DQSM?

RQ3: How can the DQSM be integrated into the architecture of data marketplaces to enhance its functionality?

To answer these questions, we adopt a problem-centered DSR approach (Peffers et al. 2008; Venable et al. 2016). We begin by conducting a structured literature review to understand the existing challenges and solutions related to data marketplaces and information asymmetry. Additionally, we engage with practitioners through interviews to gain insights into real-world issues and requirements. Based on this understanding, we derive design requirements (DRs) that inform the development of the DQSM. In the first design cycle, we conceptualize the DQSM by formulating DPs and outlining features of an artifact that conceptualizes data value asset quantification strategies, which apply ML and XAI methods. In the second design cycle, we implement a prototype of the DQSM and validate its effectiveness through technical experiments and practitioner feedback.

Our findings aim to contribute both practical solutions for data marketplaces and theoretical advancements in addressing information asymmetry. In particular, this study addresses the critical problem of information asymmetry in data marketplaces by proposing and designing a DQSM. By enabling data consumers to make informed decisions without compromising the confidentiality concerns of data providers, the DQSM has the potential to enhance the efficiency and trustworthiness of data marketplaces, fostering greater data-driven innovation and collaboration across industries.

2 Theoretical & Technical Background

2.1 The Role of Data and Barriers to Data Sharing

In today's digital economy, data has emerged as a critical strategic asset for organizations, driving innovation, enhancing decision-making, and creating competitive advantages (Wixom and Ross 2017). Companies that effectively leverage data can significantly improve their operational efficiency and market position (Spiekermann 2019). For instance, it has been a consensus for many years that data-driven decision-making boosts productivity and profitability across various industries (McAfee et al. 2012). This strategic use of data is evident in business models that rely on extensive data analytics to optimize processes, enhance customer experiences, and create new revenue streams. For example, companies like Netflix and Amazon utilize data analytics to personalize user recommendations, thereby increasing customer satisfaction and loyalty. Similarly, predictive maintenance in manufacturing and smart grids in energy management demonstrate the transformative impact of data on industrial innovation (Lee et al. 2016; Faruqui et al. 2010). Despite the clear advantages of data-driven strategies, many organizations face significant barriers to effective data sharing. These barriers can be broadly categorized into strategic, operational, technological, cultural, and regulatory challenges (Fassnacht et al. 2023).

Strategically, the lack of management commitment and integration of data sharing into corporate strategy poses a significant barrier. Often, data-sharing initiatives are project-based and driven by individual departments seeking quick wins without a top-down commitment from senior management (Abbas et al. 2021). This fragmented approach prevents organizations from fully realizing the potential of data sharing. Additionally, the absence of incentives and prospects hampers data-sharing efforts, as organizations weigh potential risks more heavily than the uncertain benefits (Gelhaar and Otto 2020). Identifying suitable use cases for data sharing is another strategic challenge, requiring transparency about existing data, creativity in developing novel services, and clarity on the benefits for all participants (Gasco-Hernandez et al. 2018; Hunke et al. 2022).

Operationally, the lack of competencies and resources required for effective data sharing presents a significant hurdle. Organizations often lack the legal, technological, and data science expertise necessary to manage data-sharing activities effectively (Zeiringer 2021). Additionally, organizations' unclear responsibilities and decision-making processes complicate the operationalization of data-sharing initiatives (Vesselkov et al. 2019). Privacy concerns also pose operational challenges, as organizations fear violating data privacy regulations when sharing and processing data (Chowdhury et al. 2020).

Technologically, barriers involve challenges related to IT infrastructure and data management. Limited data availability and accessibility due to the historical growth of disparate databases and systems impede data sharing (Bastiaansen et al. 2019). Furthermore, the lack of data processing and validation mechanisms, including data science skills and technological capabilities, hampers the ability to ensure data quality and interoperability (Choi and Kröschel 2015). Security concerns, such as unauthorized access and data breaches, require robust data security mechanisms like encryption and anonymization to foster trust in data-sharing activities (Chowdhury et al. 2020).

Cultural barriers encompass socio-cultural aspects within organizations that hinder data-sharing initiatives. A risk-averse mindset prevalent in many organizations stems from historical practices of protecting competitive information, which needs to evolve to recognize data as a strategic asset and enable unbiased decision-making (Kajüter et al. 2022). Additionally, partners' lack of trust in appropriate data usage and fears of losing control over shared data contribute to resistance against data sharing initiatives (Gelhaar and Otto 2020).

Regulatory barriers include legal and structural challenges that restrict data sharing. Compliance with legal requirements, such as data protection regulations and contract design complexities, often slows down or prevents data-sharing initiatives (Gelhaar and Otto 2020). Unclear data ownership and usage rights, especially in globally interconnected supply chains, further complicate data-sharing agreements (Pant and Yu 2018). Finally, restrictions imposed by regional, national, and international laws, along with a lack of standardization frameworks, hinder seamless data sharing across regulatory boundaries (Susha and Gil-Garcia 2019).

These barriers collectively create a complex and interdependent landscape that significantly impedes data-sharing activities among organizations. Addressing these barriers is crucial for unlocking the full potential of data as a strategic asset. In the context of data marketplaces, these challenges are particularly pertinent. Data marketplaces offer a promising solution by providing a structured environment for data sharing, addressing some of these barriers through technological and market mechanisms designed to facilitate data sharing and value quantification. The next subsection will introduce the concept of data marketplaces, exploring how they can mitigate these barriers and support data-driven innovation.

2.2 Data Marketplaces and Market Mechanisms

Online marketplaces built on digital platforms, where multiple buyers and sellers interact for trade, have transformed traditional commerce. Companies like Amazon, Alibaba, and eBay exemplify this paradigm shift (Hagiu and Wright 2015). While the terminology varies, encompassing terms like “platforms”, “e-marketplaces”, and “digital ecosystems”, the underlying concept is consistent: these environments connect multiple buyers and sellers, enabling them to conduct transactions (Constantinides et al. 2018).

This marketplace model has been extended to data trading, giving birth to *data marketplaces* (Schomm et al. 2013; Cusumano et al. 2020). Data marketplaces have emerged as pivotal platforms in the digital economy, designed to streamline the exchange of data between data providers and data consumers as an intermediary (Stahl et al. 2016; Abbas et al. 2021). These platforms provide the necessary infrastructure and market mechanisms to facilitate data transactions, addressing many barriers hindering data sharing. By democratizing access to data, data marketplaces foster innovation and create new revenue streams. Notable examples include Dawex, Snowflake Marketplace, and Caruso, which offer a wide range of datasets for various industries. The centralization and structured environment of data marketplaces enhance data discoverability and interoperability (Azcoitia and Laoutaris 2022), enabling organizations to seamlessly integrate external data into their operations and monetize their data assets effectively.

The effectiveness of data marketplaces is significantly influenced by the variety of market mechanisms they employ (Fruhwirth et al. 2020). These include different pricing strategies, such as fixed pricing, dynamic pricing models, and subscription or pay-per-use models. Fixed pricing offers a straightforward approach where data providers set a predetermined price for their datasets. Subscription models provide continuous access to datasets for a recurring fee (e.g., a tariff for full access to the marketplace or a particular data asset), which is beneficial for use cases requiring regular data updates. In contrast, usage-based pricing models adjust prices based on demand (e.g., per API calls), data quality (e.g., the granularity of time series data), and other market conditions, maximizing revenue for data providers while ensuring fair value for data consumers.

Despite these benefits, data marketplaces must address several entry barriers to ensure broad participation (Gelhaar and Otto 2020). Reducing friction for new participants involves creating user-friendly platforms that simplify the onboarding process for both data providers and consumers. Moreover, effective incentive structures are crucial; data providers need assurance of the value and security of their contributions, while data consumers require confidence in the quality and relevance of the data they purchase (Jussen et al. 2024). These incentive structures are essential to fostering a vibrant and active data marketplace Abbas et al. (2021).

To enhance the attractiveness and functionality of data marketplaces, several value-enriching features can be integrated. Data quality assessment mechanisms ensure that data consumers can trust the accuracy and reliability of the data they acquire (Abraham et al. 2023). Decision support mechanisms, such as the proposed DQSM in this article, assist consumers in evaluating the potential value of data assets before purchase (Fruhwirth et al. 2020). This addresses the Arrow Information Paradox, which highlights a fundamental issue in data transactions: potential consumers are reluctant to pay for data without first assessing its value, while data providers are unwilling to share data without compensation. This paradox results in a low willingness to pay for data, undermining traditional market mechanisms.

In the context of data marketplaces, the concept of value is inherently tied to its utility and relevance to the consumer. As (Vargo and Lusch 2018, p. 740) define, value is “an emergent, positively or negatively valenced change in the well-being or viability of a particular system/actor”. This aligns closely with the role of data assets in enhancing the ‘well-being’ of organizations through improved decision-making, operational efficiency, and innovation. Furthermore, (Vargo and Lusch 2008, p. 7) emphasize that “value is always uniquely and phenomenologically determined by the beneficiary”, underscoring the subjective and use-case-specific nature of data value. Within data marketplaces, this perspective is critical, as it frames value not as an intrinsic property of data but as a function of its potential to address the unique needs of consumers.

Drawing on foundational work in the information systems domain, value can also be linked to a data asset's ability to deliver actionable insights and support strategic goals (Grover et al. 2018; Melville et al. 2004; Schryen 2013). For instance, Grover et al. (2018) highlight the strategic importance of data in fostering innovation and creating competitive differentiation, while Melville et al. (2004) emphasize the importance of aligning IT assets, including data, with organizational objectives. These perspectives reinforce the idea that the value of data assets in a marketplace context arises from their contribution to specific outcomes, such as innovation, cost reduction, or customer experience enhancement.

Innovative solutions are necessary to overcome the challenges posed by Arrow Information Paradox. The implementation of intermediary entities that provide value assessment and selection support mechanisms can mitigate these issues. The design of a DQSM as pursued in this article, for example, enables consumers to determine the individual value of available data assets in advance, facilitating informed purchase decisions. By offering such decision support mechanisms, data marketplaces can enhance the effective use of shared data, increasing market demand and ultimately attracting more data providers Beverungen et al. (2021).

The development of these mechanisms not only addresses the Arrow Information Paradox but also enhances the overall functionality and appeal of data marketplaces. This, in turn, can drive the broader adoption of data sharing practices (Abraham et al. 2023; Fassnacht et al. 2024), unlocking the full potential of data as a strategic asset in the digital economy. By enabling consumers to make informed decisions about which data assets to purchase for their intended use case, data marketplaces can significantly enhance their effectiveness and impact. This study's focus on value quantification and selection in data marketplaces aims to provide the necessary theoretical and practical insights to advance this field, contributing to the development of more efficient and user-friendly data marketplace platforms.

2.3 Machine Learning and Explainable AI

This subsection presents an overview of crucial theoretical foundations including feature selection in ML, XAI, and active feature-value acquisition (AFA). These topics serve as grounding in related research, contributing to the rigor of the proposed solution and guiding our research design.

2.3.1 Feature Selection Techniques

ML refers to a set of methods that are used to enable computer systems to solve problems leveraging data without being explicitly programmed. A subset of ML is supervised ML, in which the aim is to train a model to generate knowledge about a set of unknown data by applying learning algorithms to a set of known data points – a core method for typical AI applications (Kühl et al. 2022). Feature selection is the process of choosing features that contribute to a target function and eliminating irrelevant ones (Blum and Langley 1997; Lessmann and Voß 2009). Benefits include alleviating the curse of dimensionality (Bellman 1966) and enhancing generalization through variance reduction (Bermingham et al. 2015).

In the feature selection domain, three primary categories of methods exist, as defined by Kumar and Minz (2014): *Filter methods* use independent measures, often statistical, to evaluate feature subsets, making them suitable for high-dimensional datasets due to their computational efficiency (Guyon and Elisseeff 2003). *Wrapper methods* examine all potential feature subsets, applying the learning algorithm to each subset. These methods, though more computationally demanding, allow for the detection of feature interactions (Kohavi and John 1997; Kozodoi et al. 2019). *Embedded methods* directly integrate with a ML model to assess a subset, often by introducing penalty terms during model training. They balance computational efficiency and predictive performance (e.g., model regularization) (Zou and Hastie 2005). Filter methods are unique as they can quantify a value based directly on the data. Conversely, wrapper and embedded methods employ a ML model. However, hybrid techniques, combining individual feature selection methods, can offer the strengths of both (Bommert et al. 2020; Saeys et al. 2007).

2.3.2 Explainable Artificial Intelligence

ML plays a significant role in a range of sectors in today's world, influencing an array of products and services (LeCun et al. 2015; Samek et al. 2017). As these technologies take on more vital tasks and decisions, there is a growing demand to comprehend their decision-making process. This has spurred the development of Explainable Artificial Intelligence (XAI), a subfield dedicated to making ML model decisions transparent and interpretable to humans (Doshi-Velez and Kim 2017; Gilpin et al. 2018). XAI methods commonly express interpretability in terms of the contribution of individual features to a prediction. Such feature importance measures not only enable the interpretation of ML models but also inform the process of feature selection, a critical step in building effective and efficient ML models (Guyon and Elisseeff 2003; Breiman 2001; Friedman 2001).

To systematically understand the field, Adadi and Berrada (2018) propose three dimensions to classify XAI methods: *Complexity* of a model has a direct impact on interpretability. Simple models are often inherently interpretable, while complex models like artificial neural networks (ANNs) or random forests are considered "black-boxes" due to their lack of inherent interpretability (Ribeiro et al. 2016a). However, this complexity often comes with improved prediction accuracy, leading to a trade-off between interpretability and model performance (Breiman 2001). *Scope of interpretability* relates to whether the interpretation targets the model as a whole (global) or specific predictions (local). Global XAI methods aim to provide an overview of a model's decision-making process by extracting important rules, while local XAI methods focus on understanding individual predictions (Yang et al. 2018). Various techniques fall under these categories, such as feature importance for global interpretability, and methods like local interpretable model-agnostic explanations (LIME) for local interpretability (Ribeiro et al. 2016b). *Level of dependency* on the ML model differentiates XAI methods based on their applicability. Model-specific methods are tailored for a specific model type, while model-agnostic methods can be applied universally, separating prediction and interpretation processes (Ribeiro et al. 2016a). This separation can be achieved through various strategies, such as training an interpretable model on the black-box model's predictions or perturbing the inputs systematically to interpret the model's reactions (Baehrens et al. 2010; Craven and Shavlik 1996).

Despite these advances, the field of XAI still faces significant challenges (Kühl 2024), including the subjective nature of interpretability and difficulties in quantifying and measuring it (Doshi-Velez and Kim 2017). The continued evolution of XAI methods will be crucial for increasing transparency and trust in ML models and for addressing ethical and regulatory concerns.

2.3.3 Active Feature Value Acquisition

AFA tackles the challenge of obtaining feature values to maximize the performance of a predictive model at minimum cost (Saar-Tsechansky 2020). This field has gained traction, particularly in scenarios where data collection or feature acquisition is costly, complex, or time-consuming. AFA strategies typically operate iteratively, deciding which instance's feature value to obtain next based on a defined criterion, such as expected model performance improvement or uncertainty reduction (Saar-Tsechansky 2020).

In the AFA domain, there are different methods regarding the timing of data acquisition and feature selection: *Induction-time AFA*: This method involves both data acquisition and feature selection during model training. New features are requested, and the model is trained on these enhanced datasets. This iterative process allows the model to learn incrementally based on the newly acquired data and the selected features (Saar-Tsechansky 2020). *Prediction-time AFA*: In this method, new features are requested and selected at the testing phase. While the model has been trained on a certain set of features, additional features may be acquired and selected during the testing phase to improve prediction accuracy (Kanani and Melville 2008). In both cases, the key decision is which features to acquire next, guided by cost, impact on model performance, or a combination of both. The chosen approach depends on the specifics of the problem and the resources available for data acquisition and model training.

Different AFA methods have been developed, ranging from heuristic-based methods (Saar-Tsechansky 2020; Kanani and Melville 2008), reinforcement learning approaches (Janisch et al. 2019; Kachuee et al. 2019), active learning inspired strategies (Li et al. 2019; Melville et al. 2005), to autoencoder-based methods (Ma et al. 2019). However, to the best of our knowledge, there is yet no approach incorporating XAI methods.

AFA approaches often have to manage varying numbers of features, as typically not all features will be acquired. Therefore, some feature values will be missing, requiring specific techniques to handle this issue. Saar-Tsechansky (2020) suggests several strategies for prediction time: Discarding instances with missing values; acquiring missing values; imputing the missing value from (a) a predictive model, (b) a distribution draw, or (c) replacement with a unique value; or utilizing reduced feature models, in which models are trained based on the available features. While strides have been made in AFA, it remains an emerging field, with ongoing research needed to fully comprehend its potential and limitations.

3 Methodology

Our research towards the design of a DQSM in a data marketplace context is guided by the DSR paradigm (March and Smith 1995; Hevner et al. 2004) and follows the three cycle view according to Hevner (2007) grounded on a relevance, rigor, and design cycle. The relevance cycle aims to identify challenges and requirements, highlighting the practical relevance of the research endeavor. In contrast, the rigor cycle ensures the incorporation of sound theories and methods from the knowledge base (Hevner 2007). Building on these practical (relevance cycle) and theoretical (rigor cycle) reflections, the design cycle aims to develop and evaluate an artifact as well as derive profound design knowledge from the design process (Hevner 2007).

Our research includes two successive design cycles (DCs) illustrated in Fig. 1, each contributing to the

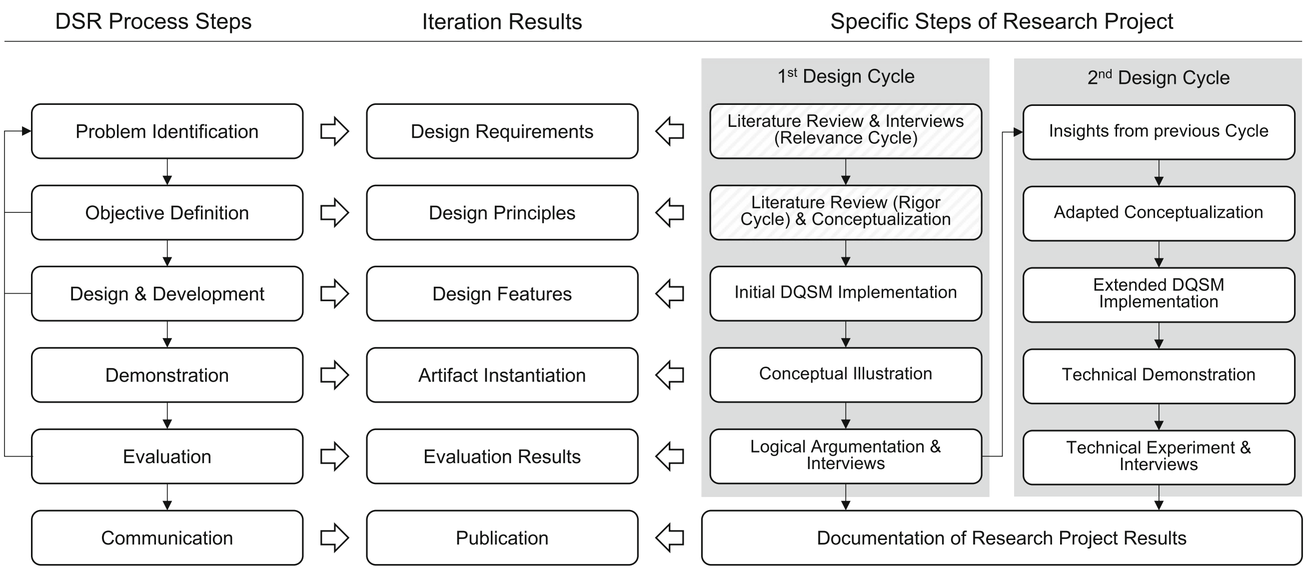


Fig. 1 Design cycles based on Peffers et al. (2008)

generation of abstract design knowledge as well as an artifact. Thus, our research contributes to the two dimensions of DSR projects according to Jones and Gregor (2007) since it provides both generalizable design knowledge and a significant impact within the field of application. Based on Peffers et al. (2008), each design cycle follows the steps of problem identification, objective definition, design and development, demonstration, and evaluation.

We base our approach on challenges related to data sharing in data marketplaces with a sensor-specific focus reported in the literature. We conduct a structured literature review according to Webster and Watson (2002), which aims to identify key challenges and inform corresponding DRs. Our literature search aims to systematically address related literature dealing with foundational components of data marketplaces to identify abstract to specific challenges using a funnel-shaped approach. Thus, we search databases from relevant disciplines (i.e., IEEE Xplore and AIS eLibrary) using the rather broad query *[(sensor OR internet of things OR data) AND (platform OR marketplace) AND (monetization OR incentive) AND (requirement OR challenge)]*. Furthermore, we conduct a forward and backward search to include other literature that the query might not capture. The initial search yielded 1948 articles, which were screened by titles and abstracts, resulting in 286 articles for full-text review. After applying inclusion and exclusion criteria focusing on peer-reviewed, online-accessible articles with an application context published between 2000 and 2023 and written in English, 53 articles were selected for detailed analysis. Data extraction was performed using a standardized form to capture key information such as application context and identified challenges.

We enrich and validate these theory-based findings by conducting and analyzing a small series of semi-structured interviews (n=4) with domain experts in data marketplaces, data monetization, sensor data, and ML (Table 1, interviewees α to δ). We apply a purposeful sampling approach based on two criteria (Palinkas et al. 2015). First, experts must have expertise in data sharing as well as foundational understanding in ML. Second, technical and business roles are included to cover heterogeneous perspectives. In the interviews, we provided two motivational and guiding use cases, asked for experiences and insights into marketplaces and mechanisms, and explored emerging trends and potential solutions. The interviews were conducted via video conferencing to allow for in-depth discussions. Furthermore, they were recorded, transcribed, and analyzed to identify recurring themes and expert opinions that complement the literature review findings.

After establishing the relevance of our DSR study, we anchor our approach in the rigor cycle by deriving DPs from kernel theories in the computer science and information systems domains, particularly focusing on feature selection techniques, XAI, and active feature-value acquisition (cf. Sect. 2.3). These kernel theories provide the justificatory knowledge that informs the design of the DQSM and ensures its grounding in established research.

Throughout the two design cycles of our DSR project, we iteratively refine these abstract DPs into concrete design features (DFs), subsequently instantiating them into prototypical artifacts. Each prototype undergoes thorough evaluation to validate and verify the design knowledge.

Table 1 Summary of interview participants. The acronyms in the column "research step" are allocated as follows: DR = Design Requirements, EE = Evaluation Episode

Interviewee	Role	Expertise	Duration (min)	Research step
α	Data Science & Artificial Intelligence Manager	Business / Technical	37	DR, EE2, EE5
β	Researcher Data Monetization	Business	51	DR, EE2, EE5
γ	Data Scientist	Technical	73	DR, EE2, EE5
δ	Researcher Data-driven Business Models, Data Marketplaces	Business	66	DR, EE2, EE5
ϵ	Data Engineer	Technical	30	EE2, EE5
ζ	Data Scientist	Technical	32	EE2, EE5
η	Researcher and Consultant for Data Strategy & Sharing	Business	77	EE2, EE5
θ	Research Group Leader & Post-Doc Digital Platforms, Industrial IoT, Ecosystem Management	Business	60	EE2, EE5
ι	Data Scientist	Business / Technical	46	EE2, EE5
κ	Senior Data Scientist	Business / Technical	46	EE2, EE5
λ	Senior Researcher Accounting	Business	52	EE2, EE5

Across the project, we conduct five distinct evaluation episodes, aligning with Venable et al. (2016)'s framework for evaluation.

Following their classification, each evaluation episode is characterized by its purpose-formative (focusing on refinement of specific aspects) or summative (assessing the artifact as a whole)-and its paradigm, distinguishing between artificial and naturalistic evaluations. While artificial evaluations explore the artifact in controlled settings, naturalistic evaluations assess it within real-world contexts, offering insights into its practical applicability. This structured approach ensures the alignment of our artifact's development with its theoretical foundations and the iterative validation of its contributions to the field.

The first design cycle develops an artifact in the form of a conceptual framework (Peffers et al. 2012) that characterizes and compares generic approaches suitable for answering RQ1. The artifact is evaluated in two separate evaluation episodes (EEs), as depicted in Fig. 2. In a first evaluation episode, the DRs and the associated DPs are reconciled, and their fulfillment is evaluated using logical argumentation. Thus, EE1 represents a formative and artificial evaluation episode according to Venable et al. (2016). Subsequently, the fulfillment of the DRs is verified through confirmatory interviews with domain experts (EE2), which represents a formative but more naturalistic evaluation. In total, we interviewed 10 experts (cf. Table 1) from different domains, covering both technical and business perspectives. Four of these interviews also helped to establish the relevance of our DSR study by validating the literature-based design requirements and guiding our technical implementation in the second design cycle. Collectively, these interviews provide practical insights and validation from experienced practitioners, ensuring that the designed conceptual artifact is aligned with real-world requirements and expert expectations.

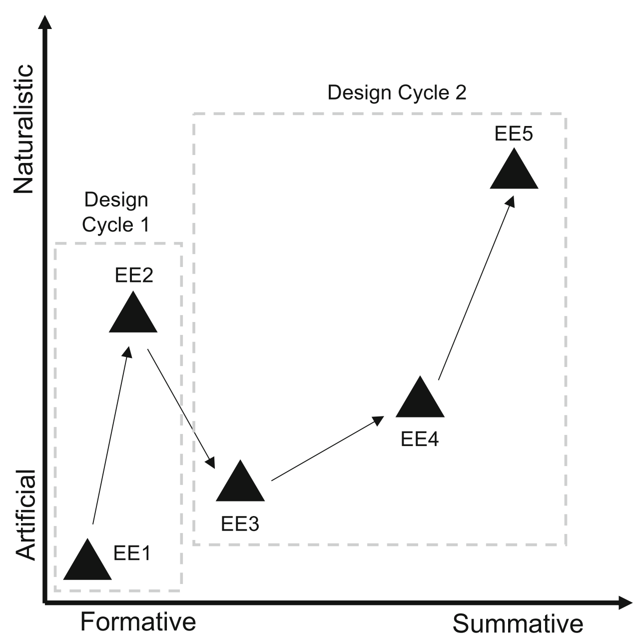


Fig. 2 Evaluation strategy comprising five subsequent evaluation episodes (EEs) based on Venable et al. (2016)

The second design cycle deepens and applies the conceptualization previously developed by implementing selected feature selection and XAI methods through a technical instantiation (Peffers et al. 2012). The artifact serves, on the one hand, to verify the technical feasibility and effectiveness (Pries-Heje et al. 2008) of the proposed conceptualization and, on the other hand, to contribute to the refinement of the design knowledge.

In the first evaluation episode (EE3), we utilized publicly available benchmark datasets commonly employed in machine learning research. These datasets were selected for their diversity in data types, task complexities, and challenges such as missing values or varying quality. This selection ensured that the artifact was assessed across a broad spectrum of use cases, providing a rigorous test of its generalizability and robustness. Detailed descriptions of these datasets, including their characteristics and challenges, are provided in the Online Appendix (see Appendix A.3, available online via http://link.springer.com). This episode aligns with a formative and artificial evaluation as per Venable et al. (2016).

In the second evaluation episode (EE4), real-world datasets were incorporated to simulate the operational conditions of data marketplaces. These datasets reflected practical challenges such as incomplete data, heterogeneous sources, and context-sensitive requirements, enabling us to evaluate the artifact's applicability in naturalistic scenarios. We used sensor-based industrial datasets for predictive maintenance to test the artifact's ability to cope with a multitude of data assets, handle time-series data and missing values.

Finally, the artifact is evaluated in terms of its usefulness through a summative and naturalistic evaluation (EE5) through expert interviews (n=11, cf. Table 1).

4 Design Cycle 1: Conceptualizing a Data Asset Value Quantification and Selection Mechanism

4.1 Problem Identification: Deriving Design Requirements

Sensor data, a subtype of internet of things (IoT) data, has unique characteristics that distinguish it from other data assets (e.g., images, text). This form of data inherently possesses a certain degree of standardization due to its intrinsic time-series nature. Therefore, it carries a temporal key, facilitating relatively straightforward integration across multiple data assets. Nevertheless, IoT data also offers several challenges, particularly in terms of completeness and quality. The occurrence of missing values is inevitable due to various factors such as device malfunction, limited precision, and failures (Klein and Lehner 2009). In addition, network communication issues and packet losses in sensor networks also contribute to data incompleteness (Wang and Liu 2011; Madria et al. 2014). Adding to the complexity, the data generated from different sensors is often collected at varying sampling rates and intervals (Lee et al. 2013). This discrepancy can further result in missing values when merging multiple data assets. The importance of this aspect is underlined by multiple concerns raised by the interviewees. Interviewee δ's query about data providers gives insights into typical gaps in data availability in an automotive data marketplace context. He questioned, "Who is actually making the data available? Are there sufficient vehicle manufacturers involved to achieve considerable coverage?". This concern indicates the need to manage situations where data might be incomplete or sparse. Furthermore, he stresses the complexity of data assets stemming from different sources by explaining, "Different types of data are available. These can be retrieved at different frequencies and in varying quality." (interviewee δ) This aspect is further amplified by interviewee β's concern about dealing with missing data. He raised, "What if 85% of the data in the dataset is missing? Do I assume what the value would be if it were complete, or do I account for the fact that it's genuinely 85% incomplete?". This highlights the necessity for the artifact to be robust enough to accommodate significant amounts of missing data, reinforcing the importance of the system's ability to handle incomplete data. In view of these issues, we propose the following DR:

DR1 (incomplete data)*: The system should be able to handle incomplete data.*

In the context of the insurance company example given in Sect. 1, DR1 addresses the potential gaps in e.g., the acquired telemetry data from vehicles due to sensor malfunctions or network issues. Thus, the system's ability to handle such inconsistencies is crucial to ensure reliable risk modeling based on driving profiles. In scientific discourse, the concept of data value is often subject to critical examination. As pointed out by Fayyad et al. (1996) and Gama (2010), raw data possesses limited intrinsic value, and its worth typically manifests only when knowledge is derived from it. This observation is further substantiated by Miller and Mork (2013) and Rowley (2007), who refer to a data value chain that illustrates the process of enriching rudimentary data (Kabadayi et al. 2005). Salminen (2018) and Gandomi and Haider (2015) argue that while big data inherently has a low-value density, it has the potential to be transformed into high value when voluminous enough. Context thus emerges as a crucial factor in augmenting the value of raw data. Sensor data, as described by Kawakami et al. (2008), is contingent upon the physical location of the sensor, a characteristic that enables context-aware applications. A relevant instance of this is user activity recognition, such as differentiating between jogging and sprinting based on the proximity to a train station or the presence of music on a user's smartphone (Campana et al. 2018).

The fact that data can be context-specific is also raised by interviewee β who argued, "If there is a context, like a certain temperature, or weather conditions, and the data is only interesting in this case, then these data are considered locally. You can consider only buying when these data are also interesting for you. From a data buyer's perspective, this is definitely exciting." This comment indicates that the system should account for the situational value of data which might change based on specific external factors, like temperature or weather conditions. Thus, we formulate the DR as follows:

DR2 (context-sensitivity)*: The system should account for the fact that the value of data is sensitive to context.*

In the insurance company case context (cf. Sect. 1), DR2 underscores the importance of context in assessing the value of sensor data. For instance, the value of vehicle telemetry data (e.g., acceleration behavior, speed, etc.) could be significantly more important to be considered when evaluated in the context of bad weather conditions (i.e., snow or heavy rainfall). This data, when viewed in a specific context, could provide more precise insights into a policyholder's driving behavior, thereby enabling more accurate risk modeling.

Both previously covered DRs can be merged into a generic design challenge addressing the data-related requirements of the system.

Value enrichment through the combination of various data assets facilitated by virtual sensors presents yet another design challenge, particularly in the context of the learning tasks required for the design of these sensors. Virtual sensors can augment the collective value of interrelated sensor data assets by relying on a sensor fusion function that integrates and aggregates specific inputs (Martin et al. 2021). This fusion function can amplify source signals either through unsupervised manual reasoning (for instance, interpreting anomalies) or via supervised ML, where a relationship between inputs and outputs is learned. Developing a fusion function hinges on the availability of labels (measured values) for a target variable intended for sensing. This output or measurement from a virtual sensor could be a continuous parameter (like temperature) or a distinct characteristic (such as defect or no defect). Thus, identifying whether the system is dealing with a classification or a regression problem is an important consideration, as highlighted by Ilyas et al. (2020). This leads to the formulation of the following DR:

DR3 (label availability/type)*: The system should be able to manage different label types.*

In the insurance company scenario (cf. Sect. 1), DR3 highlights the importance of capabilities to handle different label types in enhancing the value of sensor data. For instance, the system should be able to handle continuous parameters (e.g., speed, distance traveled) and distinct characteristics (e.g., aggressive or safe driving behavior). The vast array of fusion function methodologies in the literature underscores the necessity for flexibility in the choice of a modeling approach (Ilyas et al. 2020). These authors note the significant challenge posed by the multitude of available ML methods when developing virtual sensors and further stress that there is no universally applicable solution. Moreover, it may be viable to choose distinct modeling approaches based on specific constraints (e.g., favoring speedier training over performance when time is of the essence). Ilyas et al. (2020) also advocate for using AutoML techniques in selecting an appropriate model. Moreover, interviewee δ stresses that "the more data points there are, the more you would need a model reference. And of course, the more manufacturers are involved, the more you need something like this". These considerations underscore the need for a model-independent solution, leading us to propose the following DR:

DR4 (model independence)*: The system should be able to quantify the value of data independently or in relation to a model reference.*

In the insurance use case context (cf. Sect. 1), DR4 underscores the need for flexibility in the system to evaluate the utility of data assets in two distinct scenarios. On the one hand, it should be capable of quantifying data value independently when the insurance company does not have a pre-existing model for risk assessment. On the other hand, the system should also accommodate a reference model scenario, where the insurance company has a customized risk assessment model.

Both DR3 and DR4, on a broader note, can be merged into model-related challenges. Beyond this technical and implementation-oriented perspective, data marketplaces must provide low entry barriers to enable actors to engage actively (Guggenberger et al. 2021).

Analog to the theory of "public goods" by Samuelson (1954), a well-functioning data marketplace should be able to provide non-excludable and non-rivalrous access to all participants. Therefore, ensuring that underlying mechanisms offer tangible benefits to every stakeholder is a critical DR. These benefits can manifest in various forms, including increased access to valuable and diverse data assets (Hynes et al. 2018), improved data quality (Günther et al. 2019), efficient data acquisition and utilization (Chen et al. 2019; Anthony 2023), and potential financial gains. Agarwal et al. (2019) discusses the economic value of data and suggests that an efficient market mechanism can incentivize participants to share their data.

However, especially, ecosystems in which users mutually benefit from the provision, sharing, and use of data have specific requirements due to the unique characteristics of data as an asset (Lee et al. 2013; Agarwal et al. 2019): Firstly, data is easily replicable. Secondly, its value emerges from combination with other data, and thirdly, the resulting value varies greatly – even with the involvement of virtual sensors. Moreover, Agarwal et al. (2019) emphasizes that there is no dominant market mechanism for pricing data and matching consumers with providers. These multi-layered challenges highlight the need for incentive mechanisms facilitating that consumers pay according to their consumption and that providers deliver what was agreed upon (Misura and Zagar 2016). This is a crucial aspect, especially since the buying decision must be made ex-ante without full knowledge of the product and its potential utility (Arrow 1962; Spiekermann 2019). This is also stressed by interviewee α, who mentions that a proper evaluation of the potential value added by using a data asset must be clear before the purchase decision. The interviewees also converge on the idea of an objective intermediary who could make sound judgments about the best data for a given modeling task. By integrating these elements, the system would be capable of generating shared benefits for all parties involved.

Additionally, the participation of stakeholders is closely tied to the perceived value and trust in the system itself (interviewee δ; Sober et al. 2022). If a participant does not find value in the marketplace and the surrounding ecosystem or does not trust it, they may become less active or even choose to leave, destabilizing the entire ecosystem. It is, therefore, crucial to design the system in such a way that it provides tangible benefits and fosters trust among all stakeholders (Zuiderwijk et al. 2014). Thus, we formulate the following DR:

DR5 (stakeholder benefit)*: The system should create positive benefits for all stakeholders.*

Beyond the general incentivization aspect, the fact that ex-ante the importance of a particular asset is not known also raises the question of how to select the subjectively best asset(s) and what is an appropriate price. Interviewee γ highlights the potential of a value-based pricing model and highlights the importance of creating a cost framework that reflects the "true value derived from the data" (interviewee γ).

Agarwal et al. (2019) emphasize that consumers must make a cost-benefit trade-off to identify the best yet most affordable data assets. Lee et al. (2013) state that users are usually willing to compromise on the quality of virtual sensors if there is a high potential for savings by using less data or fewer computing resources. However, in order to make this assessment, an effective platform must provide appropriate tools that allow the consumer to make a cost-benefit judgment (Lee et al. 2013). This aspect is extended by interviewee γ extends who advocates for a "performance cost curve", which signifies the need for a quantifiable evaluation of the trade-offs between data value and its associated costs. He further comments, "as a customer, I would like to know what the data time cost to me" (interviewee γ).

Furthermore, interviewee β's viewpoint encapsulates the contrasting interests of different stakeholders, indicating, "The customer wants to see as much data as possible without having to pay too much for it. On the other hand, you have the provider, who needs to protect and monetize it." These perspectives collectively highlight the requirement for a system that provides mechanisms for an informed cost-benefit judgment to balance stakeholder interests. Thus, we formulate the following DR:

DR6 (cost-benefit advice)*: The system should advise users with a cost-benefit judgment.*

In the insurance company case context (cf. Sect. 1), DR6 suggests the platform should assist the insurance company in identifying the most beneficial and cost-effective selection of data assets for developing their risk assessment models.

Agarwal et al. (2019) stress the importance of real-time capability in a data marketplace, a reflection of the fast-paced and dynamic nature of such environments. The real-time aspect becomes crucial when dealing with the vast amounts of data that these platforms tend to handle. Given the high volume and diversity of data assets, alongside the combinatorial possibilities for data use and interpretation, these platforms necessitate efficient and scalable mechanisms to manage, process, and analyze data.

Moreover, the "combinatorial nature of data" (Agarwal et al. 2019, p. 711) implies a complex landscape of possibilities for data analysis and decision-making. As the combinations of potential data assets for utilization multiply, so does the computational complexity of processing and interpreting these combinations. This complexity can affect the speed and efficiency of the platform, which could, in turn, impact the accuracy and precision of the resulting data analysis and data-driven applications. Marketplace customers might also have varying preferences regarding the trade-off between real-time responsiveness and accuracy Lee et al. (2013). Some users might prioritize rapid results and be willing to accept a degree of inaccuracy. In contrast, others might value precision and thoroughness over speed. This variability in user preferences further complicates the balance that the system must strike between real-time performance and accuracy.

In addition, it is essential to take into account the impact of system latency on user satisfaction and engagement. As noted by Liu et al. (2010) in their study on the effects of system response time on user behavior, even minor delays can negatively affect user engagement and perceived system value. Given this, it becomes essential for data platforms to provide options for users to manage their preferences in terms of speed versus accuracy. This is also supported by interviewee δ, who asserts that "performance is definitely much more interesting. Of course, you want to make quick progress, but I think the development of a service [based on marketplace data] takes time anyway. Whether it can be delivered a few weeks later or not doesn't matter. Much more interesting is when the service is then tested, how quickly can I retrieve the data from the marketplace, or can I even stream it in real time? Without this streaming, some use cases don't work at all. So this topic, when do I get the data, I wouldn't see as really relevant. The quality must, of course, be right". This opinion illustrates the trade-off between real-time access and data quality, underlining the need for a system that can allow users to customize their preferences accordingly.

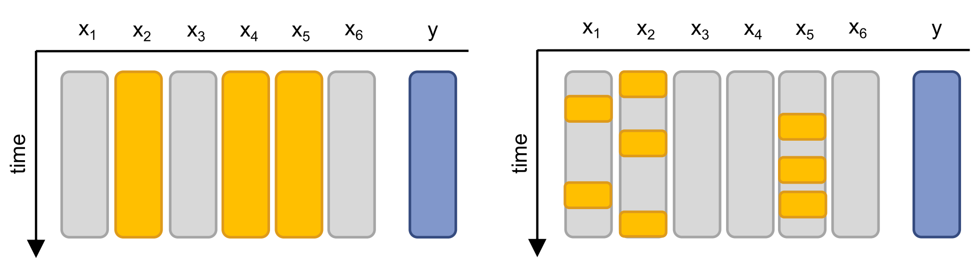
Similarly, interviewee γ reinforces this aspect by stating, "I feel, [response time] wouldn't be that important to me. As a customer now. I would have the feeling okay I am applying now, and looking at my case I would like to know what the dataset would cost me, and whether they report back in half an hour or next week, it's pretty much the same to me. That's not so relevant for me". This perspective reflects a clear preference for the quality of the data over its real-time accessibility. Both opinions emphasize the need for a system that lets users define their priorities, as well as whether they lean more towards immediate data availability or data accuracy. Therefore, we derive the following requirement:

DR7 (real-time vs. accuracy)*: The system should enable the user to define their preferences in terms of real-time versus accuracy.*

4.2 Objective Definition: Deriving Design Principles

This section derives DPs from the justificatory knowledge to set clear objectives for our artifact design. In line with the data-related DRs stipulated in DR1 and DR2, as well as DR7, the proposed artifact should accommodate the assessment of data while considering the individual context. As depicted in Fig. 3, this assessment can be distinguished into a *static* (i.e., without context) and a *dynamic* assessment (i.e., with context). We consider our automobile insurance use case introduced in Sect. 1 for clarity. The company is building a risk model that factors individual driving behaviors based on various vehicle metrics. A static assessment would collectively assess the impact of a particular data asset (e.g., speed) without regard for individual circumstances. In contrast, a dynamic assessment would assess the effect of speed on a case-by-case basis, factoring in context-specific data like weather conditions, therefore providing a much more nuanced picture (DR2). This concept closely aligns with the 'scope of interpretability' criterion as posited by Adadi and Berrada (2018) as a way to differentiate various XAI methods. Drawing upon this differentiation outlined in Sect. 2.3, we can distinguish between global and local XAI methods. Further, we can combine these methods with strategies from the AFA domain to address missing values (DR1), as proposed by Saar-Tsechansky (2020).

Though inspired by XAI methods, our groups are, in essence, feature selection strategies, hence necessitating distinct terminology. Thus, we propose 'static assessment methods' for global inspections, which return a single contribution score for a given feature across all instances, and 'dynamic assessment methods' for local inspections, which return multiple contribution scores for a feature, accounting for varying significance in different contexts. These deliberations lead us to formulate:



(a) Static assessment

(b) Dynamic assessment

Fig. 3 This figure demonstrates the process of modeling a virtual sensor (y) using a total of six sensor data assets ($x_1..x_6$). In a static assessment scenario **a** x_2, x_4, and x_5 would offer a sufficiently good global contribution to the objective function y. In contrast, a dynamic assessment **b** would find that local feature-value combinations of x_1, x_2, and x_5, depending on the context, contribute equally well to the objective function y but require fewer data assets

DP1 (context-awareness)*: Provide the system with adaptable mechanisms that assess data assets either in a static or dynamic manner.*

Moreover, the system, according to DR3 and DR4, should be able to handle various ML models and different label types. To realize this objective, the artifact needs to be designed flexibly to support individual data examination and to incorporate usage of a ML model, as specified by DR4. Furthermore, it must effectively handle different label types to fulfill DR3.

In relation to the insurance company case, the system should offer flexibility to operate with or without a user-provided pre-existing risk assessment model. This capability ensures that the insurance company can use the platform effectively regardless of whether they have a proprietary risk model available. In addition to these functional considerations, some non-functional ones arise from usability-related challenges: Taking DR7 and DR1 into account, the different selection approaches should also account for different response time requirements (DR7) as well as data availability (DR1) requirements: More specifically, this means that approaches should be available for situations where a short run time is expected or only a limited amount of data can be used for inspection (Misura and Zagar 2016).

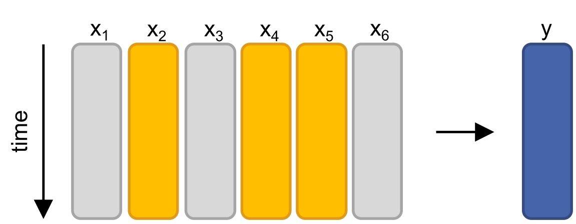
As illustrated in Fig. 4, we differentiate between data-based and model-based assessment mechanisms: *data-based assessment mechanisms* should score the individual contribution of each data asset (i.e., feature) to a given target based on just the data itself, while *model-based assessment mechanisms* should return a contribution score, taking into account a pre-existing user-defined ML model which describes a relationship between data assets and a given target. The different characteristics of these two strategies make it possible to fulfill the above-mentioned requirements in different ways: While data-based mechanisms are less complex approaches that do not necessarily require a ML model, the complexity and thus the computing time for model-based mechanisms is typically higher (Molnar 2019). Thus, we formulate the following DP:

DP2 (level of dependency)*: Provide the system with adaptable mechanisms to assess data assets using either data-based or model-based strategies.*

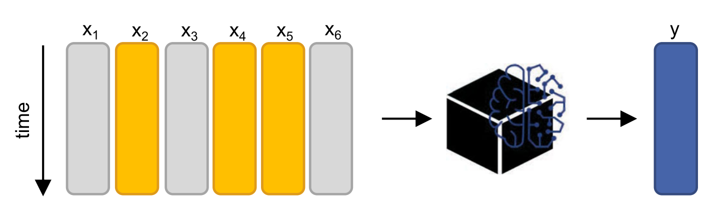
In order to make the platform attractive for providers and consumers to participate in the platform, it should provide an incentive for all parties involved (DR5). As described by Van Alstyne et al. (2016), a platform ecosystem consists of a platform owner/provider, producers, and consumers. In order to fulfill DR5, the platform should provide value to all these players:

With regard to the *platform owner*, the artifact should provide value through matchmaking between data providers (acting as producers) and data customers (acting as consumers). To facilitate this matchmaking, the platform should provide value in two ways: First, through assessing all available data and providing a consumer with information about the most valuable features for his particular use case. Second, it should acknowledge the dynamic nature of data, considering changes over time and varying environmental states. For instance, the driving behavior during winter (e.g., reduced speed, increased breaking) could significantly impact the risk model and thus be a valuable feature during these months but might not hold the same relevance in summer months. This capacity for dynamic adaptability enhances the utility of the platform for its users.

In terms of adding value for the *data provider*, it can be accomplished by enhancing data sales while preserving data privacy and reducing marketing efforts. This system tackles the Arrow Information Paradox (Arrow 1962) by ensuring customer satisfaction without requiring pre-purchase data sharing. This reduces entry barriers and alleviates the need for detailed data asset marketing, as important attributes can be derived and presented without disclosing raw data. Therefore, it increases the provider's ability to sell data while maintaining privacy and reducing overhead costs.



(a) Data-based assessment



(b) Model-based assessment

Fig. 4 This figure illustrates the creation of a virtual sensor (y) utilizing six sensor signal time series ($x_1..x_6$). In a static assessment scenario **a** we select the most influential data assets without referencing any pre-existing model. In contrast, a model-based approach **b** incorporates a user-provided model into the inspection process

For the *data consumer*, it also provides value through decreasing uncertainty and ensuring the promised contribution of the purchased data to a given ML model. Furthermore, this reduces the cost of training a ML model and fosters reliability (Arrow 1962), which in turn provides the advantage of better budget control for the consumer (DR6). This depiction offers a minimal approach to understanding platform ecosystems and may overlook certain stakeholders, such as e.g., app providers, regulators, etc., while also focusing more on the roles rather than individual actors. Finally, it is to be noted that it is not possible to clearly indicate a single party favored by certain aspects due to the inherently connected structure of two-sided markets within platforms, where more than one party benefits if the value is created for one of the parties (Van Alstyne et al. 2016).

In conclusion, these aspects can be summarized with regard to their common objective: To enrich the value of data for the benefit of all roles involved in the data sharing platform. The enrichment of data can be interpreted in the context of the DIKW hierarchy presented by Rowley (2007). This model suggests the enrichment of data in order to create information, knowledge, and wisdom. Transferred to this case, the value assessment of data can provide such an enrichment, as detailed above. Thus, we formulate another DP:

***DP3 (value enhancement)**: Provide the system with capabilities to enhance the value of data assets for all stakeholders.*

Furthermore, the system should actively assist customers in determining a reasonable budget (DR6) for acquiring data. This is crucial as there is an interplay between cost and model performance when training a ML model with data sourced from a data sharing platform. However, purchasing additional data doesn't necessarily guarantee improved model performance. The optimal use of a given budget lies in acquiring the right data rather than more data. Hence, it is vital to consider each feature's incremental contribution to a ML model against its associated cost to make effective and beneficial purchasing decisions (DR5). For instance, consider a dataset with two correlated features – a model could potentially extract significant information from only one of those features, reducing the necessity for the second one (James et al. 2013). The decision to incur an additional cost for the second feature is contingent on the specific use case and should be left to the user. The system's role should be to facilitate informed decision-making by providing relevant comparisons:

***DP4 (decision support)**: Provide the system with features that support informed decision-making.*

4.3 Design and Development: Mapping Design Principles to Design Features

In this section, the previously introduced DPs are mapped to DFs. The relationship between DRs, DPs, and DFs is shown in Fig. 5.

As discussed in Sect. 2.3, current research highlights the potential of XAI and feature selection techniques for quantifying and assessing data value. These techniques fundamentally differ: XAI methods base their value estimation on the decision-making process within a ML model

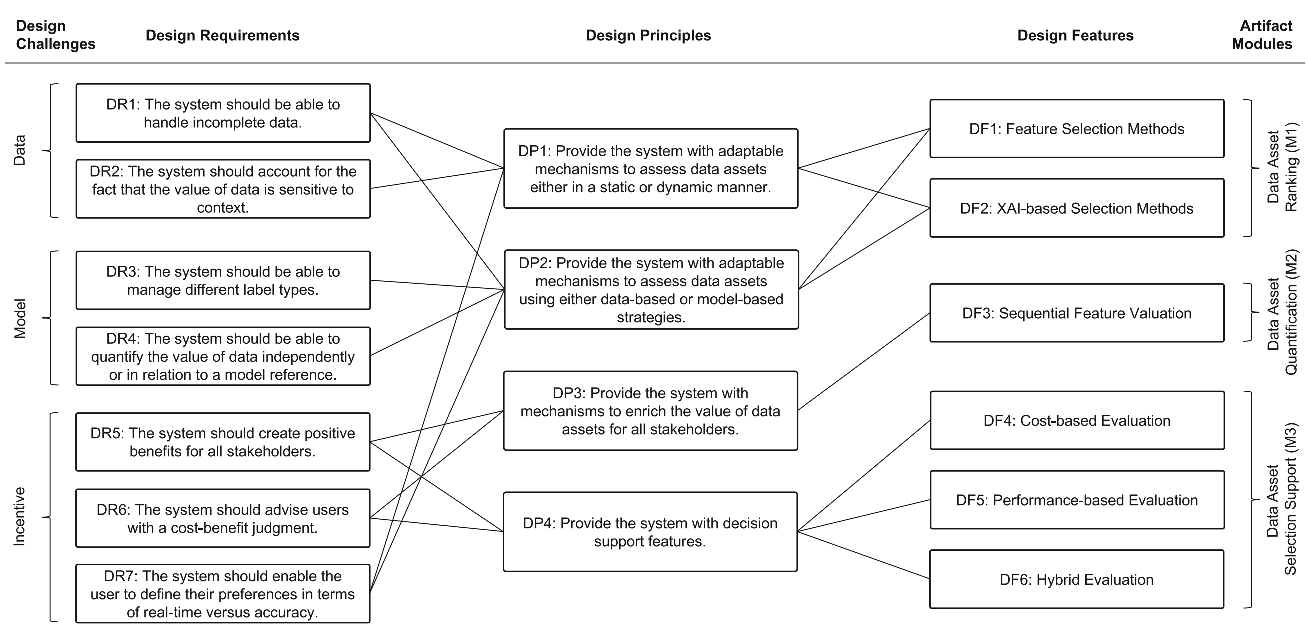


Fig. 5 Mapping from design requirements to design features

(Arrieta et al. 2020), whereas feature selection techniques offer a robust alternative when missing labels or other conditions hinder the training of a ML model (Bermingham et al. 2015). Furthermore, in general, feature selection techniques are less computationally expensive than XAI techniques. This has an impact on the run time and also on the required computing power. Thus, feature selection methods serve as complementary tools for fulfilling different aspects of the DPs. For instance, DP1 calls for a distinction between static and dynamic assessment, while DP2 stipulates that the system should be capable of data assessment based on ML models. As depicted in Fig. 7, this results in four possible combinations, each of which can benefit from either XAI or feature selection methods. Consequently, we propose two DFs: data asset value assessment leveraging feature selection techniques (DF1) and assessment utilizing XAI techniques (DF2). These DFs aid in developing diverse data asset selection strategies in line with other DPs, while using different method groups allows us to capitalize on their unique strengths.

DP3 mandates that the system should enrich a data asset's value with additional context and meaning in alignment with the DIKW pyramid's (Ackoff 1989) systematic approach. Though DF1 and DF2 form the technical foundation for quantitatively evaluating individual data assets, converting this quantitative assessment into a meaningful format for the consumer facilitates the calculation of the marginal performance contribution of each data asset. This requires considering the data assets' value not in isolation but in relation to other selected assets in the constructed dataset.

It is crucial to note that the total value of a dataset isn't simply the sum of individual feature values – effects such as correlation can reduce the total value drastically (Arrieta et al. 2020). Therefore, we outline DF3, tasked with converting these evaluations. DF3 uses the estimated individual value of each data asset, identifies the most promising one yet to be selected (i.e., the data asset with the highest importance score), and evaluates the marginal performance gain caused by the considered asset, given that a set of other assets were already selected. This proposed heuristic greedy search approach considers feature interactions based on the intuition of sequential feature selection (SFS) (Pudil et al. 1994) but does not take all possible combinations into account. However, such an exhaustive approach would be far more expensive and is therefore not realistic for wide data (Ferri et al. 1994).

DP4 implies that decision support should help users decide not only which specific features to purchase but also how many to buy. We propose two considerations: the level of performance achievable with a set budget and the budget needed to attain a desired performance level. These considerations build the foundation for the implemented DFs: a cost-based evaluation (DF4), which focuses on the performance achievable within a given budget, and a performance-based evaluation (DF5), which seeks to understand the budget necessary to reach a desired performance level.

As noted earlier, there is a relationship between costs and performance. Decision support should consider this interconnected structure by facilitating a balance between the two target variables. We refer to this third type of evaluation as hybrid evaluation (DF6). The elbow criterion (Thorndike 1953) serves as a useful analogy for this hybrid evaluation. In the k-means clustering algorithm, the optimal number of clusters is typically determined by plotting the sum of squared distances over the number of clusters chosen. The "elbow", or the point where the curve is tangent to a curve with a slope of -1, represents the optimal number of clusters. Similarly, an optimal budget can be identified from a plot of performance over cost.

4.4 Demonstration: Artifact Illustration

Based on the DFs, the initial artifact is divided into three modules depicted in Fig. 6 on page 17:

The *Data Asset Ranking Module (M1)* module implements a data asset importance quantification strategy,

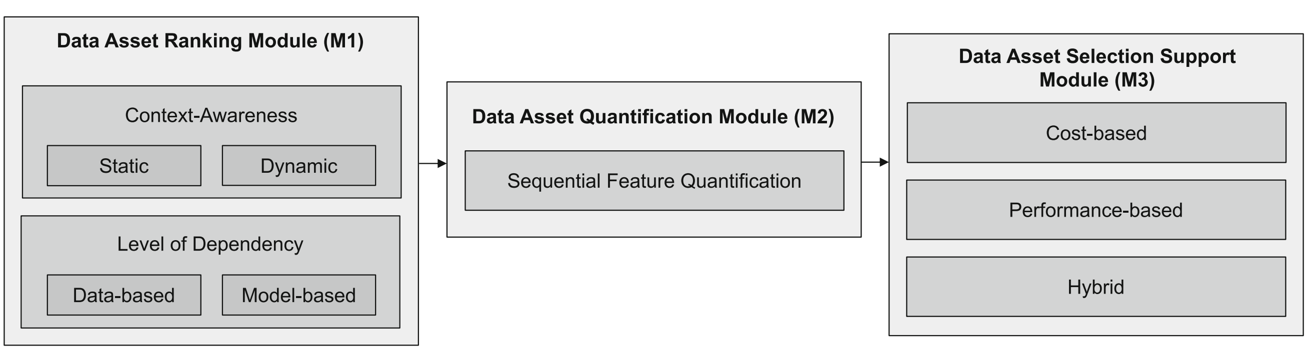


Fig. 6 Data asset value quantification and selection mechanism (DQSM) artifact modules

encompassing DF1 and DF2, which stem from DP1 and DP2. Given the inherent nested structure of these principles and features, a 2x2 matrix, as visualized in Fig. 7 on page 18, can serve as a representation of this module. This matrix categorizes the distinct characteristics that the DPs embody, with each quadrant of the matrix corresponding to suitable techniques – feature selection and both global and local XAI methods. This matrix acts as a foundational framework to classify and detail various data asset selection strategies underpinned by DF1 and DF2. These strategies aim to quantify a particular data asset's contribution, and thus the value, to a given problem.

The *Data Asset Quantification Module (M2)* is designed to augment M1. While M1 lays the methodological groundwork for quantifying the predictive power of an available data asset to a given target, M2 integrates these methods, translating the importance estimation into actionable insights by deriving the available data assets according to their marginal contribution to a given objective function. Given that this process operates sequentially based on previously quantified scores, the underlying process is denoted as 'sequential feature quantification' (DF3).

These insights are subsequently processed through the *Data Asset Selection Support Module (M3)*. The core objective of this module is to relay the outcomes of the previous modules effectively. It centers on presenting the information derived from M2. Thus, the module's output (DF4-DF6) aims to offer the marketplace customer the decision support stipulated in DP4.

4.5 Interim Evaluation (EE1, EE2)

As articulated by Hevner (2007), the relevance cycle dictates the application context, offering "acceptance criteria for the ultimate evaluation" (Hevner 2007, p. 3). Relevance informs solution acceptance, thus, we examine how well the initial artifact meets these challenges.

Kuechler and Vaishnavi (2008) further posit a second design cycle for an extensive evaluation of the first one. In this context, we gauge the artifact's effectiveness against the challenges, as per the framework by Petter et al. (2010), while the second design cycle substantiates the artifact's feasibility, setting the stage for subsequent evaluations. For this purpose, the individual challenges formulated in Sect. 4.1 are revisited. In the first evaluation episode, we use logical arguments (Peffers et al. 2012) to address the DRs individually, assessing how well the artifact aligns with them and its potential as a solution.

The collection of data, for example in the IoT context, faces several obstacles, notably missing data through several issues (DR1) (Klein and Lehner 2009). The artifact leverages methods from the AFA field to address this issue, particularly through the dynamic approaches within data asset ranking strategies, which accommodate different

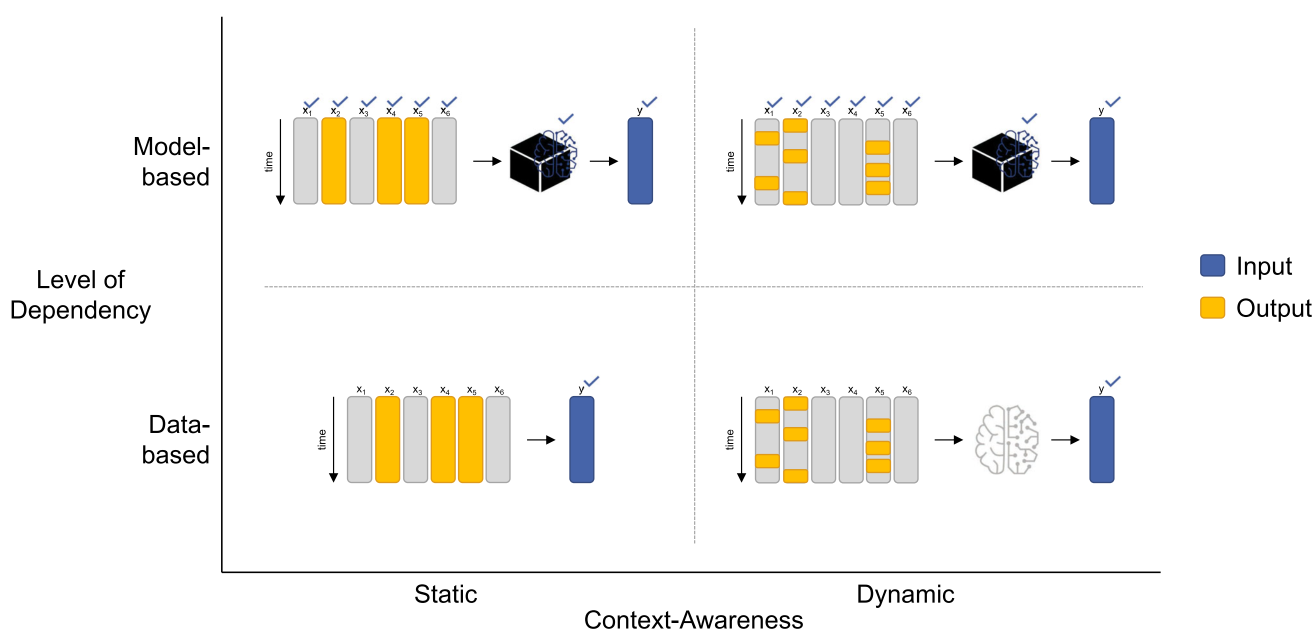


Fig. 7 Scenario matrix for data asset ranking. The four quadrants of the matrix address different scenarios in terms of context awareness and level of dependency, which determines the selection of suitable methods. The horizontal bars $x_1 \ldots x_5$ represent data assets (i.e., features in a dataset), whereas y corresponds to the target. Blue elements indicate inputs required for the DQSM in the different scenarios (historical target data and, in the model-based case, additionally, a black-box model and its feature mapping). Yellow elements show the output of the mechanism, i.e. the selected feature (values)

features per instance, thus allowing flexible data asset selection when a desired feature is absent. The artifact also considers the challenge of transforming data into contextually meaningful (DR2) information (Salminen 2018; Gandomi and Haider 2015; Kawakami et al. 2008). It incorporates the ability to quantify data value on varying granularity levels, including static and dynamic aspects, thus facilitating the assessment of the impact of different granularity levels. Data label type (DR3) and model availability (DR4) present another challenge. Thus, our artifact is designed to handle data-based and model-based tasks, as well as regression and classification problems (James et al. 2013) with a set of different methods, thereby providing a suitable technique for all identified scenarios. The artifact's data asset quantification module addresses the challenge of incentivizing participation (DR5) of all involved roles through ensuring platform functionality (Agarwal et al. 2019; Misura and Zagar 2016; Spiekermann 2019). These are generally addressed by the ability to objectively quantify the value of a data asset without transferring or revealing the data. Furthermore, an informed cost-benefit judgment (DR6) on data asset selection is central to customers of a data marketplace (Agarwal et al. 2019). Our proposed data asset selection support module enables customers to make decisions aligned with their circumstances. Fast response time (DR7) is essential (Agarwal et al. 2019), though its importance may vary depending on the application. Therefore, the artifact distinguishes between data and model-based approaches, leveraging the general tendency of feature selection methods to have low runtimes compared to XAI methods and providing the user the option to choose accordingly.

In a second evaluation episode, we build on interviews with domain experts to evaluate the fulfillment of the DRs. The experts (cf. Table 1) were asked to assess whether the proposed artifact met the specified DRs. We used a 5-point Likert scale with options ranging from 'strongly disagree' (1) to 'strongly agree' (5). This approach provides quantitative data on the fulfillment of each DR, allowing for an objective measure of the artifact's success. Table 2 aggregates the responses and provides an average and standard deviation for each DR. The results show overall a strong agreement that the artifact fulfills all given DRs, with DR4 (model independence) having the lowest score.

In addition to the quantitative results, the interviews provided valuable qualitative feedback on several DRs: Referring to DR4, interviewee α raises a concern that architectural modifications of the model would be necessary when new data assets are introduced. He notes, "At the moment you decide to introduce more GPS data, you have a new architecture; you need to rebuild and train it again" (interviewee α). This feedback relates specifically to cases where a model is provided by the user, but there may not be directly compatible data assets available. Thus, the feedback highlights the system's need for flexibility and adaptability to accommodate evolving data landscapes. Additionally, interviewee β expressed uncertainty about the definition of value in DR4, indicating, "The value is still hard to grasp. Is it a value between 0 and 1, or is it trying to quantify it in Euros, what is the goal?" This statement underlines the necessity of clear and specific value measures. Lastly, for DR6, interviewee γ remarked that real-time aspects of the system may not be a crucial factor for customers. To some extent, this relativizes the importance of DR7 compared to the other DRs. These qualitative insights help to supplement the quantitative evaluation and provide a nuanced understanding of aspects to revise in the subsequent design.

Table 2 Results of DR fulfillment evaluation (n=11)

Design requirement	Avg.	SD
DR1 (incomplete data)	4.5	0.8
DR2 (context-sensitivity)	4.4	0.9
DR3 (label type)	4.5	1.0
DR4 (model independence)	4.1	1.6
DR5 (stakeholder benefit)	4.8	0.4
DR6 (cost-benefit advice)	4.3	1.3
DR7 (real-time vs. accuracy)	4.1	1.3

5 Design Cycle 2: Prototype Development

The objective of the second design cycle is to refine and substantiate the findings of the first design cycle. For this purpose, we aim to implement the initial artifact in Python. This implementation serves two purposes: First, to assess the feasibility of the initial artifact. Second, to further solidify the artifact and also enable a quantitative evaluation. For this purpose, this section is divided into three parts: First, for each module of the artifact, we determine how the rather theoretical conceptualization from the first design cycle's artifact can be translated into a practically implementable concept. Subsequently, we outline the overall implementation process. Finally, we present the outcomes of three different EEs.

5.1 Objective Definition

In order to make the initial artifact, designed in the first design cycle more tangible, the second design cycle aims to further specify technical and implementation-related details. Therefore, we build on the individual modules of the initial artifact introduced in Sect. 4.4.

As specified in Sect. 4.4, the purpose of the *data asset ranking* module (M1) is to provide the methodological basis to quantify the predictive power of a given data asset in relation to a given target. As illustrated in Fig. 7, we propose four different strategies to quantify the importance of a given data asset by building on various methods from the XAI, feature selection, and AFA domains. As a basis, M1 requires access to the entire feature set representing a variety of data assets offered on a marketplace. Moreover, the user (i.e., potential customer) has to provide historical target data to enable the DQSM to offer a tailored selection of useful data assets. Furthermore, in the case of model-based scenarios, the user needs to provide a trained black-box[4] model or application programming interface access to a deployed model. The output of this module consists of rank scores indicating the predictive power of a data asset. These scores are unique per feature in static assessment scenarios and unique per feature value in dynamic selection scenarios (as highlighted via yellow elements in Fig. 3).

The remaining part of this section aims to specify the individual strategies associated with each quadrant of the matrix illustrated in Fig. 7. The concrete implementations presented are chosen based on their extensive use and acceptance in both academia and practice. In addition, their selection was influenced by the existence of high-quality, open-source Python implementations that make them widely accessible and applicable.

Static assessment mechanisms aim to quantify the overall relative contribution of a given individual data asset (i.e., feature) to an objective function, regardless of the context. In accordance with DP2, this can be achieved either in a data-based manner, using feature selection mechanisms (DF1) or based on a ML model, utilizing XAI methods (DF2). Therefore, the two static selection quadrants of the matrix leverage different implementations: *Static data-based assessment mechanisms* assess the contribution of a given data asset based on feature selection mechanisms, as introduced in Sect. 2.3. Well established implementations are: variance-based, analysis of variance (ANOVA), mutual information, and SFS (Bommert et al. 2020). Further details on these approaches are described in Online Appendix A.1. All these methods can be implemented using the scikit-learn and NumPy libraries in Python. Unlike the data-based approaches, *model-based approaches* inspect a pre-existing (black-box) ML model in order to quantify the importance of data assets. Therefore, model-agnostic XAI methods can be applied to this model. As introduced in Sect. 2.3, we can distinguish between global and local XAI methods, based on the criterion 'scope of interpretability' (Adadi and Berrada 2018). While global XAI methods seem to be intuitively suitable to static selection procedures, it should be noted that local methods can be used as well, through aggregating the scores of individual instances in order to receive an overall score (Lundberg and Lee 2017).

Global XAI methods we use include: permutation importance, partial dependence plots (PDP), accumulated local effects (ALE) (Apley and Zhu 2020), and ELI5 model inspection. In addition, local XAI methods we utilize include shapley additive explanations (SHAP) and LIME. Further information on the methods is provided in Online Appendix A.2. While some of the presented methods (e.g., permutation importance, ELI5, SHAP, LIME) return numeric feature importance scores already, other approaches are more focused on illustrating the model behavior graphically (e.g., PDP, ALE). These methods have been adapted to the needs of this study by extracting numerical values to quantify individual features' importance. These approaches are based on the range and standard deviation of the quantile scores, which would otherwise be plotted.

In accordance with DP1, *dynamic assessment mechanisms* should provide the means to inspect a model locally and return different importance scores for a given feature for different contexts. In order to achieve this, it is necessary to enable the inspection method to differentiate these different contexts. This can be achieved through the application of local XAI methods (i.e., SHAP, LIME) to assess the marginal contribution of each individual feature value to the objective function. Further, acknowledging that local XAI methodologies can be computationally demanding in particular on large datasets, we suggest an alternate approach based on context-sensitive segmentation of the data into several partitions. This allows for a global assessment within each partition. In our study, k-Means-based clustering is employed for such segmentation, although other methodologies may also be suitable. With this data partitioning strategy, static evaluation mechanisms can also be utilized for dynamic assessment. However, the context sensitivity is restricted compared to local XAI techniques and depends on the number of clusters. Nonetheless, it is a less computationally expensive alternative. Consequently, in terms of *dynamic data-based assessment*, we can apply local XAI methods to a surrogate model or employ feature selection or even global XAI strategies paired with a clustering-based partitioning approach. Yet, it should be noted that for large datasets, local XAI methodologies may not be an ideal solution due to their computational complexity. Thus, we recommend a two-step approach that initially employs a computationally

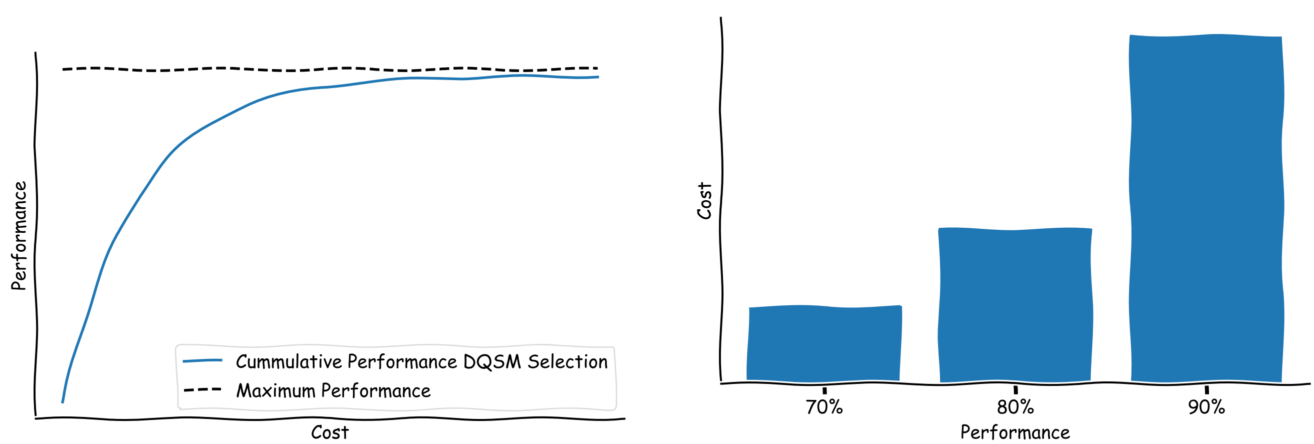
[4] Such a user-provided model may be a black-box model in the sense that the exact functionality remains hidden from the platform. Only the inputs must be specified, and a mapping to data assets available on the platform must be given.

efficient pre-selection via feature selection methods on contextual clusters, followed by a refined ranking via local XAI.

Dynamic model-based assessment strategies use a user-provided ML model and thus do not depend on all available data assets on the marketplace. Instead, they rely on a customer-specific subset used to train the model. This approach facilitates the use of local XAI. Nevertheless, the cluster-based methodology remains a feasible alternative for this scenario as well. Based on these rank scores derived by M1, the *data asset quantification* module (M2) aims to determine the marginal contribution of a data asset to a given objective function. The underlying 'sequential feature quantification' process proposed in Sect. 4.4 takes the data asset rank scores returned from M1 as input and feeds the data assets according to the sequence given by the ranking to an evaluation model. In the case of model-based scenarios, the user-provided model acts as an evaluation model, whereas in the case of data-based scenarios, we employ a surrogate model. Non-selected feature values are masked through a perturbation strategy to maintain the overall distribution of the features. The next step involves scoring the model based on the masked dataset. Afterward, the cumulative contribution scores are scaled to a range between 0 and 1. This scaled score is subsequently weighted according to the maximum possible model performance. The final outputs of M2 are the marginal contributions of each data asset to the model performance, providing a valuable measure for decision-making in the context of data asset acquisition and utilization enabled through the *data asset selection support* module (M3). In DC1, we outlined various evaluation methods to facilitate a user's decision-making process when training a ML model (DF4-DF6). To illustrate the foundational concepts, we offer idealized depictions in Fig. 8a, b.

The cost-based evaluation primarily aims to ascertain the quantity and choice of features needed to develop an optimally performing model within a specified budget (DF4). This is premised on the notion that the additional performance gain contributed by each successive feature diminishes as more features are already utilized. Thus, a user can determine the expected performance corresponding to a particular budget. This degressive relationship is illustrated in Fig. 8a. The figure illustrates the expected model performance (vertical axis) for a given budget (horizontal axis). The sequential feature evaluation approach approximates performance based on the entire dataset, and this level of performance is achieved once all features are used.

The blue trajectory depicted in Fig. 8a schematically illustrates the scenario assuming equal prices for all data assets, a simplification often observed in real-world cases. Concurrently, our approach facilitates the applicability of a value-based pricing strategy. This is possible since M2 quantifies the marginal contribution of each asset towards a prediction, which aligns with the customer's marginal benefit concerning a given modeling task. With a value-based pricing strategy in place – implemented by the platform – the trajectory would follow a diagonal course. Consequently, the performance gain per added data asset would equate to the additional cost. Even with free pricing set by the providers, the performance curve is beneficial to the customer as a decision-making tool. This is because the curve is likely to display multiple inflection points, resulting from the over- and under-proportional prices relative to the improvement in performance. Moreover, the performance-based evaluation (DF5) aims to quantify the cost associated with achieving a specific model performance level. To select the appropriate set of data assets, users can refer to a plot, such as the one sketched in



(a) cost-based

(b) performance-based

Fig. 8 Qualitative depiction of cost-based and performance-based evaluation

Fig. 8b. This bar chart demonstrates the costs incurred to achieve a set performance level. Lastly, the hybrid optimization approach (DF6) strives to strike a balance between cost and performance, based on Fig. 8a. Mirroring the elbow criterion used to choose the number of clusters in k-means clustering, a tangent can help identify the most cost-effective performance benefit. By adjusting the tangent's gradient, user-specific requirements (performance vs cost) can be weighted appropriately, if necessary.

5.2 Design, Development, and Demonstration

In this step, we aim to extend the artifact developed in DC1 to materialize the abstract conceptualization into a fully functional software artifact. The primary objective is to create a system that not only fulfills the outlined objectives but is also designed in a flexible, scalable, and user-friendly manner. We choose Python as the implementation language due to its widespread usage and strong support for ML tasks. Python's vast ecosystem of libraries, particularly in the realm of XAI, offers a foundation of already implemented methods that expedite the development process and minimize the likelihood of bugs through reliable open-source tools (Bissyandé et al. 2013). One of the critical aspects of our technical implementation is to establish an architecture that aligns with the conceptual model yet is flexible enough to accommodate any future modifications or extensions. To this end, we employed an object-oriented approach to design the system's structure, which provides modularity and enhances the maintainability of the system. Our implementation structure reflects the DQSM framework depicted in the conceptual design phase. It provides functionalities to load and handle various datasets, to assess the value of individual data assets (i.e., features) using the full range of data asset ranking and value quantification strategies presented in Sect. 5.1, and to translate these evaluations into actionable insights using the data asset selection support plots.

5.3 Ex-post Evaluation (EE3, EE4, EE5)

The second design cycle aims to substantiate the conceptualized design knowledge gained in DC1 through an instantiation (Peffers et al. 2012). The design of our artifact serves two primary objectives: First, as suggested by Pries-Heje et al. (2008), it aims to validate the practical feasibility and effectiveness of the proposed conceptualization. Second, it facilitates the refinement and advancement of the developed design knowledge. In a third evaluation episode (EE3), we utilize widely known and publicly accessible datasets to conduct a technical experiment that assesses the artifact's performance. This formative and artificial evaluation, as classified by Venable et al. (2016), aims to assess the general feasibility of the instantiation as well as the technical performance of the system. Thus, we conduct experiments on ten well-established public datasets for ML tasks. As detailed in Table 3, the datasets represent different sizes and both classification and regression labels. Since meaningful and wide IoT datasets are not widely available as well as usually very specific and hardly intuitive, we intentionally use more general data in this evaluation episode, although the scope of our study is focused on IoT data due to their better interoperability. Despite this expansion of the scope, we do not see any limitation as the shown framework is, in principle, applicable to all data, but in practice, such non-IoT datasets can hardly be compiled from different sources as mapping is a hurdle due to the lack of a common key.

For the purpose of illustrating the basic functionality in this article, we employ the widely known Boston Housing dataset (Fig. 9). This dataset includes 14 variables and 506 instances, describing the average house prices in the Boston area (Harrison and Rubinfeld 1978). For dynamic data-based scenarios, we implement a random forest model with default parameters. We choose this model since it is among the most widely used (Buitinck et al. 2013). For model-based scenarios, we leverage XGBoost (Chen 2016), which generally provides marginally superior performance on our chosen datasets. Thus, this model is likely to be a preferred choice for customers in the given cases and, furthermore, can be considered as a black-box model. However, our solution offers a broad variety of models included since we build on a modular basis which works with all scikit-learn models, as well as CatBoost (Dorogush et al. 2018) or Keras (Chollet 2015), among others. We make use of the

Table 3 Summary of datasets

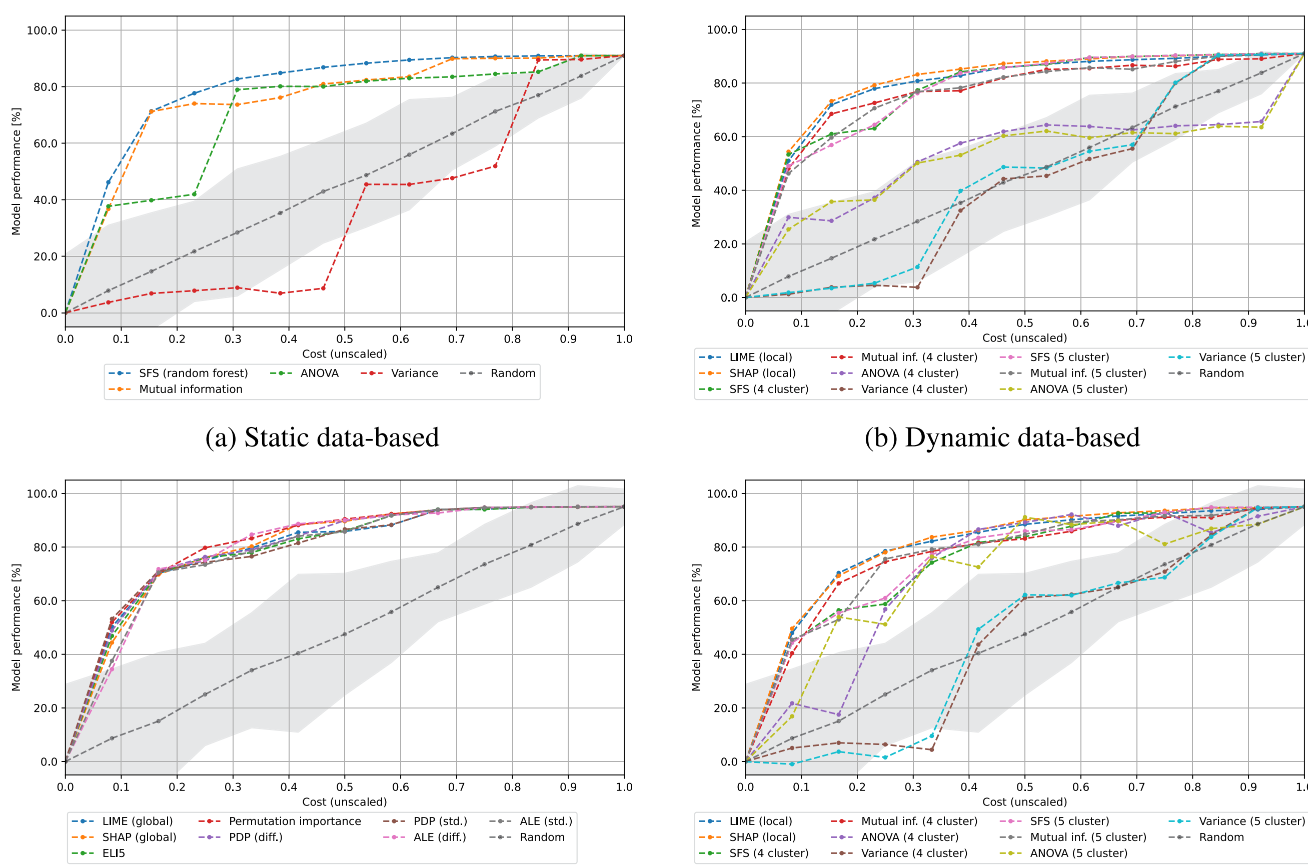
Dataset	# Features	# Instances	Task
Adult (Becker and Kohavi 1996)	14 (106^5)	48842	Classification
Auto MPG (Quinlan 1993)	7	398	Regression
Boston Housing (Harrison and Rubinfeld 1978)	14	506	Regression
Breast Cancer Wisconsin (Wolberg etal. 1995)	30	569	Classification
California Housing (Pace and Barry 1997)	8	20640	Regression
Diabetes (Kahn 1994)	10	442	Regression
Gas (CO2) (Kaya et al. 2019)	11	36733	Regression
Gas (NOx) (Kaya et al. 2019)	11	36733	Regression
Heart Disease (Janosi et al. 1988)	14 (26^5)	303	Classification
Wine (Aeberhard and Forina 1991)	13	178	Classification

R2 score for regression tasks and the F1 score for classification tasks due to their broad acceptance and interpretability in the field.

To illustrate the outcome of our experiments, we use both cost-based and performance-based evaluation methods to compare various data selection strategies. Furthermore, we report the performance of a random feature selection as a benchmark. The model performance in all our experiments is evaluated by averaging 50 repetitions of feature permutations as a masking strategy for non-selected data assets.

Figure 15a (Online Appendix) depicts the performance of various static data-based strategies evaluated in a cost-based manner. Except for the variance-based method, which fails completely to capture any relevant data assets in an early acquisition phase, all other strategies surpass the average performance of random selection significantly. The SFS-based strategy clearly outperforms ANOVA and the mutual information method over the entire cost range between 0 and 0.7, except for one characteristic point where SFS and mutual information methods are equally good. This point marks the selection of the second feature, the combination of which both methods agree on.

Figure 15b (Online Appendix) presents the cost-based evaluation results for the dynamic data-based strategies. SHAP in particular stands out here with a very stable and high performance. LIME's performance is only slightly weaker over the entire course. Mutual information also shows quite high performance in the 4- and 5-cluster variants. In contrast, in this example, the variance-based strategies are again at most at the random choice level, and the ANOVA-based cluster approaches also only perform



(a) Static data-based

(b) Dynamic data-based

(c) Static model-based

(d) Dynamic model-based

Fig. 9 Results of the cost-based evaluation performed on the Boston Housing (Harrison and Rubinfeld 1978) dataset. For the static data-based scenario (**a**), we report sequential feature selection (SFS) based on a random forest surrogate model, analysis of variance (ANOVA), variance, and mutual information-based selection strategies. For the dynamic data-based scenario (**b**), we report local interpretable modelagnostic explanations (LIME) and shapley additive explanations (SHAP) utilizing a random forest surrogate model alongside the cluster-based SFS, ANOVA, variance, and mutual information methods. For the static model-based scenario (**c**), we report LIME and SHAP as global approaches, permutation importance, PDP, ALE, and ELI5. Finally, for the dynamic model-based scenario (**d**), we report LIME and SHAP as local approaches alongside the cluster-based SFS, ANOVA, variance, and mutual information methods. The black trajectory represents the mean score of a random-choice-based selection accompanied by a tolerance band (i.e., $[\mu - 3\sigma; \mu + 3\sigma]$), which is derived from 50 repetitions of the process

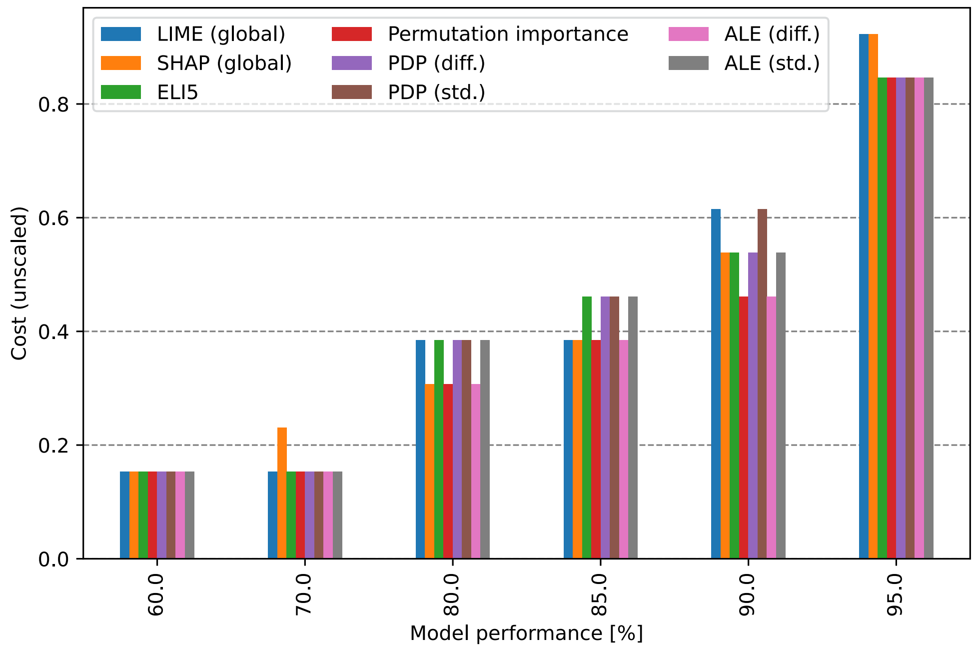


Fig. 10 Static model-based performance-based evaluation

well in the initial phase. In the static model-based scenario (Fig. 15c Online Appendix), we use an XGBoost model as a basis, which results in a slight increase in the maximum model performance. Moreover, it can be observed that all approaches achieve competitive results. In this case, permutation importance performs best, especially in the middle part, ahead of SHAP, PDP (diff.), and LIME. However, the differences are comparatively small. In the dynamic model-based case illustrated in Fig. 15d (Online Appendix), we observe again clear differences. As in the dynamic data-based scenario, SHAP achieves the highest values, closely followed by LIME. Mutual information again performs well in the cluster variants. ANOVA and variance-based selections again achieve results worse than random.

The performance-based evaluation illustrated in Fig. 10 provides consumers with a simple visualization to compare the costs incurred for certain performance levels at a glance. We leverage the performance-based plot in this article to compare several approaches and see slight differences. In an actual implementation of DQSM in a data marketplace, however, we argue that only one strategy per scenario must be implemented.

Figure 11 shows a comparison of the best-performing methods across all four quadrants of the scenario matrix. The plot indicates that regardless of the underlying concrete assessment method, the artifact can achieve comparably good results. Moreover, dynamic methods (i.e., SHAP (local) and LIME (local)), in particular, show a slight performance gain with relatively few selected data assets due to the context sensitivity in corresponding context-rich datasets. However, this advantage decreases with more assets added.

Subsequently, in a fourth evaluation episode (EE4), we utilize real-world sensor data, aiming to mimic the nature of data that could be encountered in an actual IoT data marketplace. This more naturalistic evaluation provides valuable insights into the performance and applicability of the presented artifact in a real-world scenario. In the considered industrial use case, the goal is to develop a condition monitoring service for hydraulic seals. Thus, we aim to develop a virtual sensor to distinguish between normal operation and various failure modes. Therefore, we utilize data gathered on a test bench. The data sources include a variety of sensors placed in the surrounding of the sealing element. These sensors measure diverse parameters, such as temperatures, pressures, motion, friction, vibration, and acoustics. The high-frequency sensor signals (i.e., acoustics and vibration) are preprocessed into about 510 distinct time series, each matching the sampling rate of the remaining sensors. Additional environmental conditions like room temperature and humidity, not in the direct proximity of the sealing element, are also measured, leading to an overall total of 545 sensor features.[5]

The user-provided model we use for model-based scenarios is a deep artificial neural network. As a surrogate model, we employ a random forest model with 500 estimators but no additional hyperparameter tuning. In this case, the data originates from a single test bench, but in a field application, there could be different measurements

[5] After transforming the categorical variables to binary features employing one-hot-encoding (Hancock and Khoshgoftaar 2020).

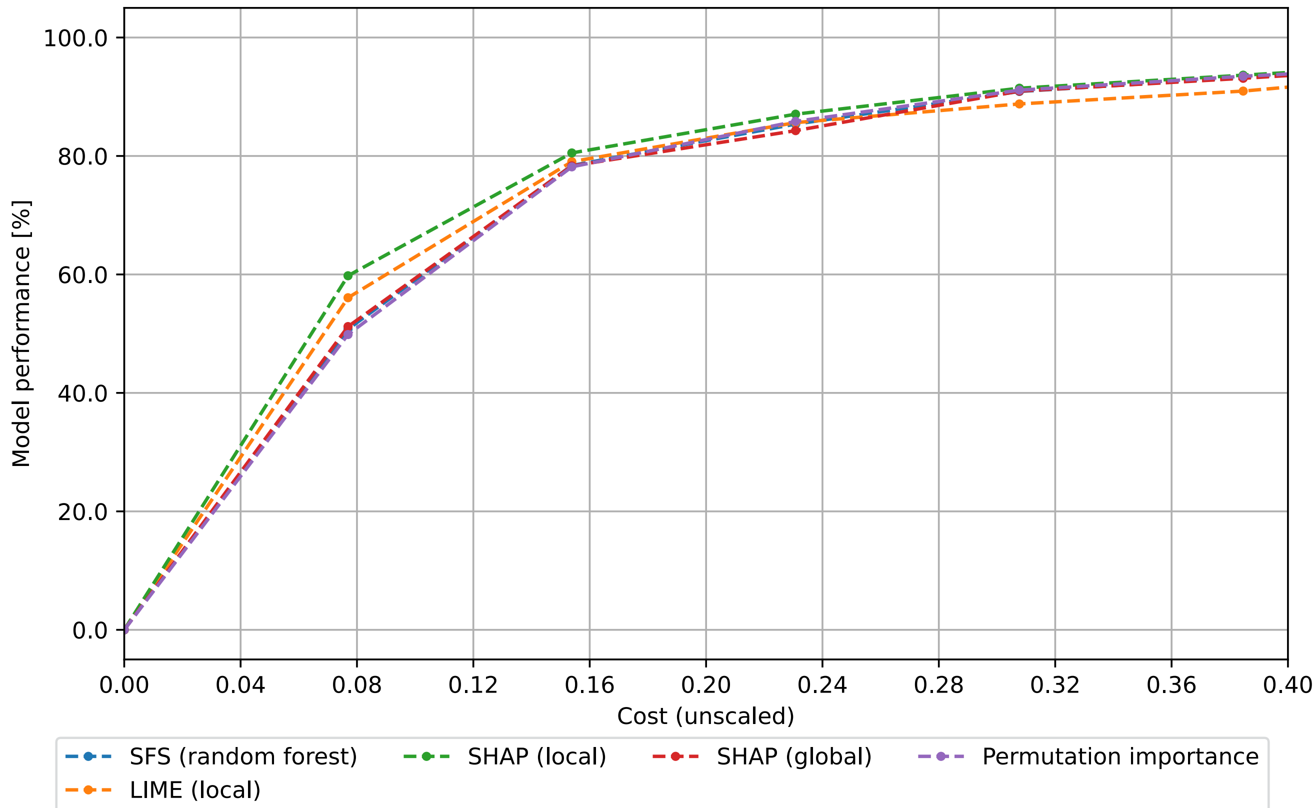


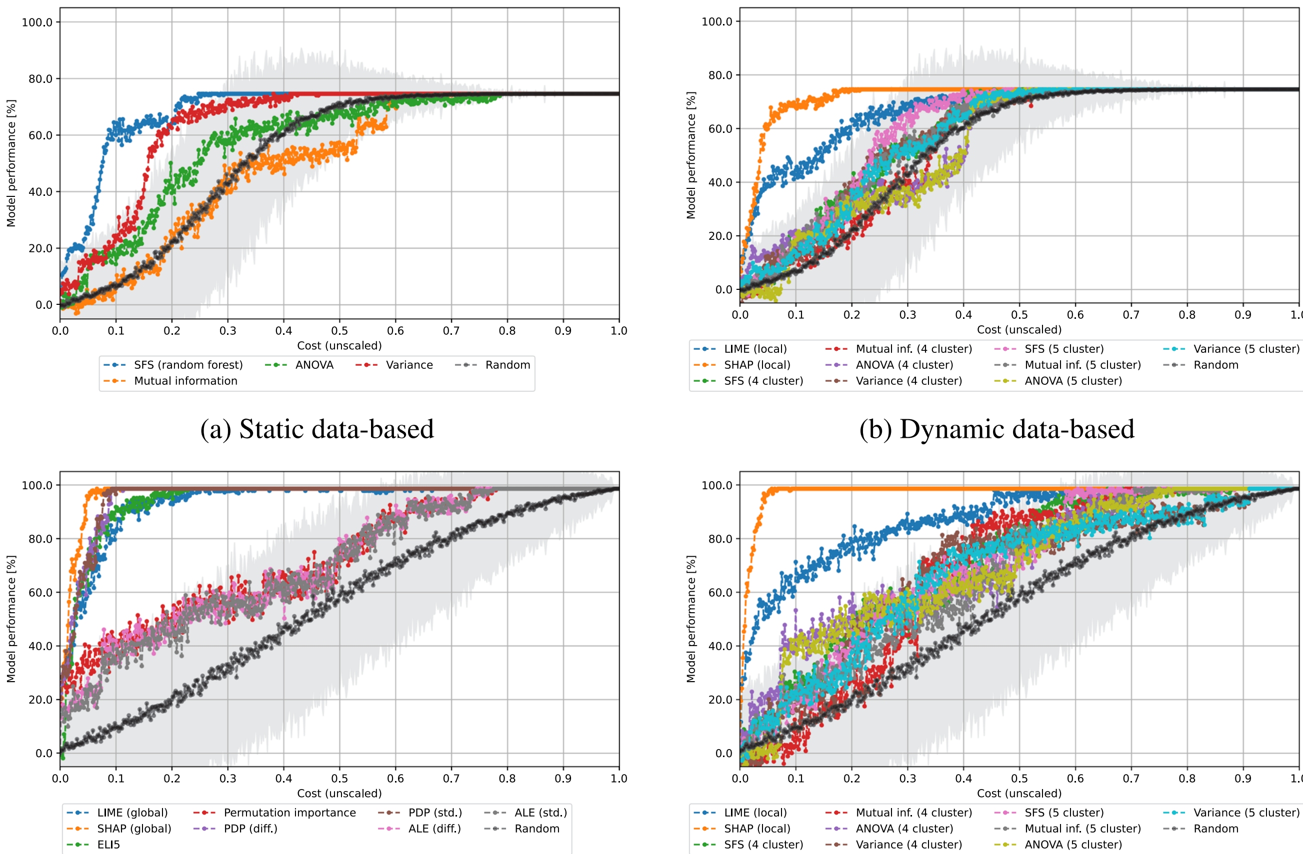
Fig. 11 Comparison of best-performing methods across all four scenarios via cost-based evaluation

from different actors involved in the production and operation of an end-use application (e.g., a wind turbine). Therefore, the use presented case is a suitable example to evaluate our proposed DQSM artifact. Figure 12 illustrates the cost-based evaluation plots for the four scenarios. The static data-based methods are weaker compared to the remaining three quadrants, but in line with the general trend, SFS outperforms the other methods, especially early on. The dynamic data-based selection shows quite good results of SHAP. But also LIME performs well. In the static model-based case, even all methods except for ELI5, ALE (std.), and permutation importance perform almost equally well. In the dynamic model-based case, the results are overall comparatively weaker at the front. Only SHAP performs very well and yields a very steep gradient towards maximum performance. Overall, it is also evident from a significantly wider tolerance band in the model-based scenarios that the user-provided model is much more sensitive to the permutation strategy we employ for masking. In this case, it might make sense to focus more on data-based or static model-based selection.

This evaluation again confirms that effective data asset assessment and selection is supported by our artifact in all four strategies.

In EE3 and EE4, we critically assess the applicability and adaptability of our approach through extensive testing of the artifact on 11 diverse datasets. The results across these datasets are notably consistent, as illustrated in Table 4. For better comparability, we calculate the area under the curve (AUC) for the cost-based evaluation curves. The curves are scaled to a range between 0 (worst performance) to 1 (best performance) to exclude model-related performance differences. Regardless of the dataset dimension, modeling task, and scenarios, our approach demonstrates its capacity to reliably quantify the importance of data assets. Overall, for static data-based scenarios, SFS turns out to be consistently strong, whereas for static model-based scenarios SHAP and permutation importance and for dynamic scenarios SHAP emerge as top-performing methods across all datasets.

Furthermore, the compatibility with different ML models enhances the utility and applicability in practice. Therefore, we conduct tests involving various ML models, ranging from simple tree-based and linear regression models to complex ANNs. This broad compatibility ensures that our approach can adapt to different predictive modeling tasks while maintaining consistent performance. Nevertheless, we see differences in performance depending on the selected model. Figure 5 shows that on average a



(a) Static data-based

(b) Dynamic data-based

(c) Static model-based

(d) Dynamic model-based

Fig. 12 Results of cost-based evaluation on our industrial seal condition monitoring dataset. The black trajectory represents the mean score of a random-choice-based selection accompanied by a tolerance band (i.e., $[\mu - 3\sigma; \mu + 3\sigma]$), which is derived from 50 repetitions of the process

random forest achieves slightly better AUC scores than XGBoost, the chosen representation of a user-specific model. We also see similar performance differences with other models but do not go into more detail about the differences in this study. Despite the absence of hyperparameter tuning in our experiments, our approach yielded robust results. This implies that the proposed DQSM is not overly sensitive to the hyperparameters and offers a solid baseline performance. However, this also indicates an additional potential for further performance enhancement, particularly with a focus on the so far, compared to, e.g., SHAP, less effective approaches based on clustering techniques.

In terms of computational efficiency, we observe a trade-off between performance and computational expense. As depicted in Fig. 13, the methods with superior performance (i.e., SFS, SHAP, permutation importance, and ALE) across all test cases tend to be more computationally demanding, though still within manageable bounds. All our experiments were conducted on a limited-resource cloud computing instance, implying that the computational requirements of our approach are reasonable.

Finally, in the fifth evaluation episode (EE5), the artifact's perceived usefulness and the user's attitude towards using the artifact, central constructs in the technology acceptance model (TAM), are evaluated through a summative and naturalistic evaluation. This assessment provides a user-centric perspective and verifies user acceptance and satisfaction, thereby aligning the artifact more closely with real-world user expectations and needs.

Table 5 illustrates the evaluation results of six items related to the TAM concepts 'perceived usefulness' and 'attitude toward use' (Davis 1989). We used a 5-point Likert scale with options ranging from 'strongly disagree' (1) to 'strongly agree' (5). Users gave high scores for the perceived usefulness of the DQSM, with an average score of 4.8 and 4.1 on its potential to increase analytical performance and productivity and a 4.6 score for its ability to assist in selecting the right data for their tasks. The attitude towards using DQSM was also rated positively, with a 4.8

Table 4 Comparison of ranking methods across datasets. The values illustrate the AUC for scaled cost-based evaluation curves. High values indicate an effective selection of data assets, causing a steep increase in performance through early selected data assets

Scenario	Ranking method	Datasets										
		Adult	Auto MPG	Boston	Breast	California	Diabetes	Gas (CO2)	Gas (NOx)	Heart	Wine	Industry
Static data-based	ANOVA	0.7332	0.5759	0.7168	0.8616	0.3957	0.6559	0.6597	0.5570	0.7301	0.7569	0.7778
	Mutual information	0.6759	0.5847	0.7733	0.8623	0.6537	0.6423	0.6547	0.6633	0.7043	0.7524	0.6850
	SFS (random forest)	**0.8281**	0.5649	**0.8130**	**0.8693**	**0.6579**	**0.6756**	**0.6615**	**0.7118**	**0.7727**	**0.7670**	**0.9244**
	Variance	0.8230	**0.6119**	0.3593	0.7952	0.3024	0.4168	0.3825	0.4333	0.6516	0.6243	0.8615
Dynamic data-based	LIME (local)	0.5675	0.6498	0.8080	0.8665	0.6510	0.6671	0.6707	0.6843	0.7608	0.7619	0.8930
	SHAP (local)	**0.8951**	**0.6563**	**0.8216**	**0.9072**	**0.6894**	**0.7142**	**0.6754**	**0.7317**	**0.8616**	**0.8074**	**0.9616**
	SFS (4 cluster)	0.7383	0.5895	0.7917	0.7967	0.6273	0.6585	0.3717	0.7045	0.7493	0.6592	0.7783
	SFS (5 cluster)	0.7345	0.5915	0.7851	0.7912	0.6080	0.6597	0.3274	0.6872	0.7518	0.6441	0.8019
	ANOVA (4 cluster)	0.4572	0.5400	0.5464	0.7913	0.3963	0.3538	0.4409	0.6455	0.6225	0.6611	0.7433
	ANOVA (5 cluster)	0.4635	0.5358	0.5325	0.7748	0.3480	0.4232	0.4173	0.6455	0.6173	0.6588	0.7301
	Mutual inf. (4 cluster)	0.6221	0.5898	0.7787	0.7634	0.6378	0.5227	0.4328	0.6624	0.6967	0.6599	0.7361
	Mutual inf. (5 cluster)	0.6094	0.5862	0.7718	0.7694	0.6327	0.5279	0.4306	0.6366	0.6587	0.6040	0.7715
	Variance (4 cluster)	0.8174	0.6049	0.4309	0.7909	0.3466	0.4647	0.2892	0.4140	0.6764	0.6312	0.7788
	Variance (5 cluster)	0.8160	0.6077	0.4530	0.7971	0.3357	0.4403	0.2453	0.4227	0.6768	0.6379	0.7688
Static model-based	LIME (global)	0.8440	0.6904	0.7771	0.8687	0.6573	0.6146	0.5358	0.6692	0.7992	0.7966	0.9488
	SHAP (global)	**0.9333**	**0.6930**	0.7826	0.8790	0.6576	0.6450	0.5624	0.6688	**0.8020**	**0.8269**	**0.9795**
	ELI5	0.9197	0.6304	0.7748	0.8551	0.5636	0.5991	0.5533	0.6649	0.6508	0.7316	0.9554
	ALE (diff.)	0.9091	0.6876	0.7783	0.5920	0.6602	0.6314	0.4934	0.6695	0.7817	0.5135	0.7189
	ALE (std.)	0.9077	0.6886	0.7669	0.5873	0.6610	0.6293	0.5371	0.6643	0.7884	0.5143	0.7100
	PDP (diff.)	0.9079	0.6867	0.7828	0.8704	0.6608	0.6260	0.6253	0.6690	0.7992	0.8213	0.9663
	PDP (std.)	0.9090	0.6724	0.7749	**0.8793**	0.6207	0.6160	0.5971	0.6677	0.8012	0.8214	0.9665
	Permutation importance	0.9301	0.6786	**0.7948**	0.8587	**0.6620**	**0.6477**	**0.6354**	**0.6726**	0.7976	0.8240	0.7312
Dynamic model-based	LIME (local)	0.8245	0.6820	0.7788	0.8551	0.6591	0.6239	0.4205	0.6419	0.7800	0.7517	0.8839
	SHAP (local)	**0.9535**	**0.6923**	**0.7863**	**0.8877**	**0.6745**	**0.7050**	**0.5426**	**0.6918**	**0.8505**	**0.8465**	**0.9843**
	SFS (4 cluster)	0.8732	0.6462	0.7362	0.7998	0.4883	0.6428	0.3493	0.6635	0.7470	0.6383	0.7204
	SFS (5 cluster)	0.8736	0.6613	0.7394	0.7502	0.5119	0.6364	0.2500	0.6314	0.7505	0.6027	0.7028
	ANOVA (4 cluster)	0.6497	0.4673	0.6840	0.7768	0.4544	0.3607	0.3333	0.6193	0.6073	0.6348	0.7206
	ANOVA (5 cluster)	0.6554	0.4932	0.6832	0.7504	0.3753	0.4359	0.3290	0.6198	0.5976	0.5936	0.7040
	Mutual inf. (4 cluster)	0.8487	0.5527	0.7483	0.7026	0.5960	0.5320	0.2587	0.6065	0.6377	0.6387	0.6845
	Mutual inf. (5 cluster)	0.8533	0.5802	0.7476	0.7005	0.5940	0.5185	0.2880	0.5553	0.6291	0.5663	0.6736
	Variance (4 cluster)	0.9114	0.6417	0.4471	0.7627	0.4118	0.4540	0.3295	0.4618	0.6532	0.6663	0.6850
	Variance (5 cluster)	0.9109	0.6458	0.4443	0.7564	0.4006	0.4300	0.2590	0.4507	0.6391	0.6499	0.6815

Bold values highlight the highest value per scenario dataset combination. BoldUnderline values highlight the best AUC among a dataset

Table 5 Results of user acceptance and satisfaction evaluation (n=11)

Item	TAM Construct	Avg.	SD
Using DQSM would increase my analytical performance	Perceived usefulness	4.8	0.4
Using DQSM would increase my analytical productivity	Perceived usefulness	4.1	1.2
Using DQSM could support me in selecting the right data for my tasks	Perceived usefulness	4.6	0.5
I think that using DQSM is a good idea	Attitude	4.9	0.3
I think that using DQSM is beneficial to me	Attitude	4.8	0.4
I have a positive perception about using DQSM	Attitude	4.5	0.9

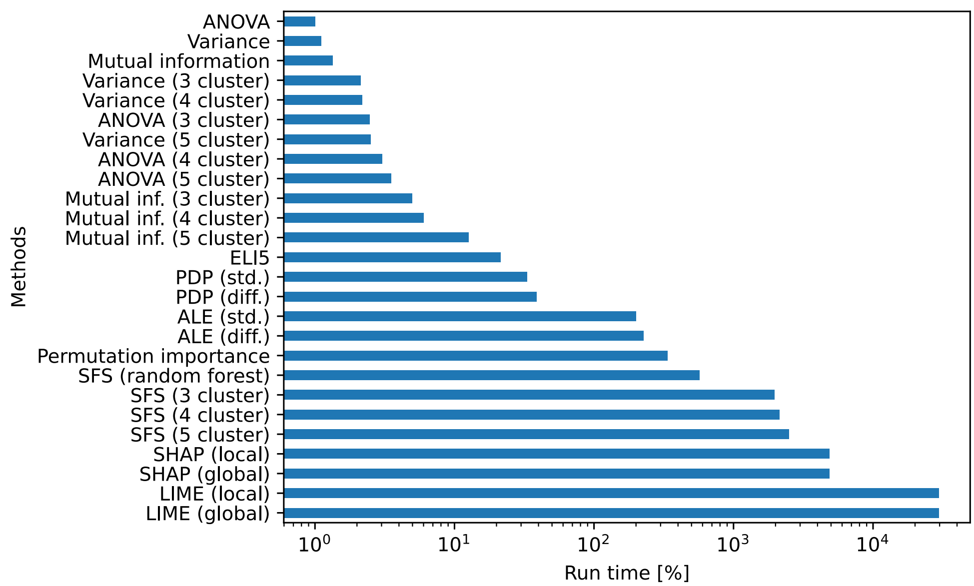


Fig. 13 Comparison of average run times of presented methods. The graph serves as an overview of the proportional differences between the methods across our tested datasets. However, a generalizable relation cannot be derived because different methods have different dependencies on the selected model, the number of features, and/or the number of instances

average score for the interviewees thinking that using DQSM is both a good idea as well as beneficial. The overall positive perception about using DQSM had an average score of 4.5. The results underscore the artifact's potential value and strong user acceptance, verifying its relevance and usability in practical scenarios.

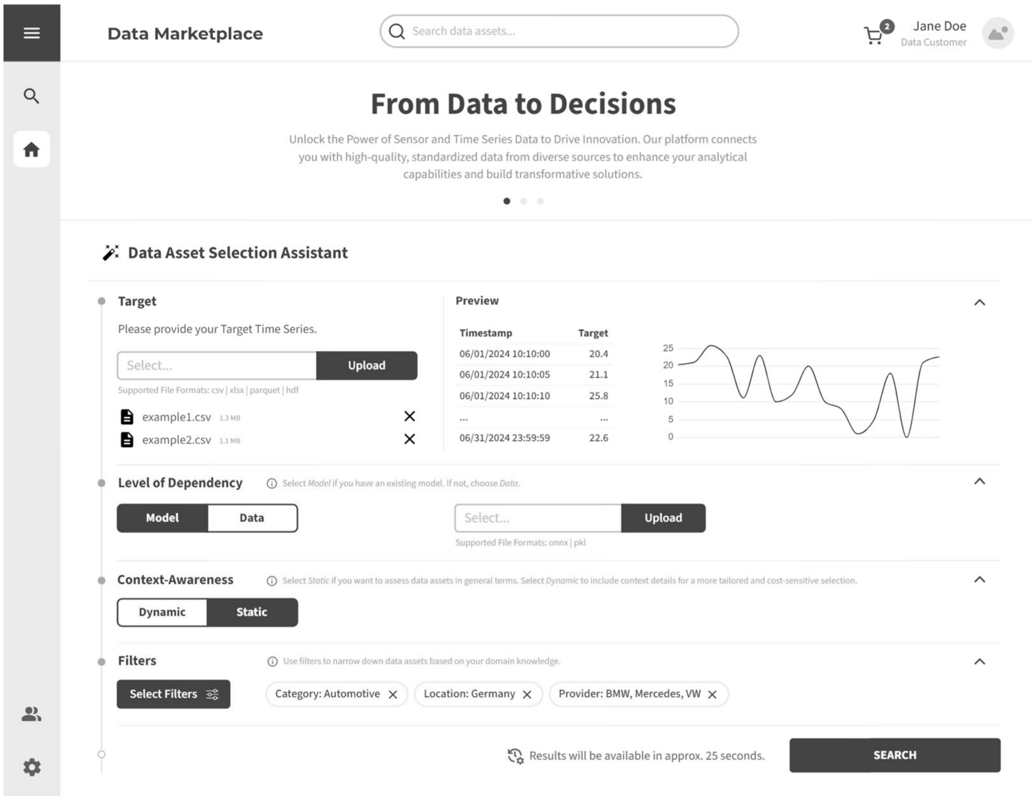
6 Discussion

Building on our findings, we now outline the necessary steps for deploying the DQSM in an operational setting (Sect. 6.1), discuss its contributions to both theory and practice (Sect. 6.2), position our work within the broader IS domain (Sect. 6.3), and reflect on its limitations to inform future research directions (Sect. 6.4).

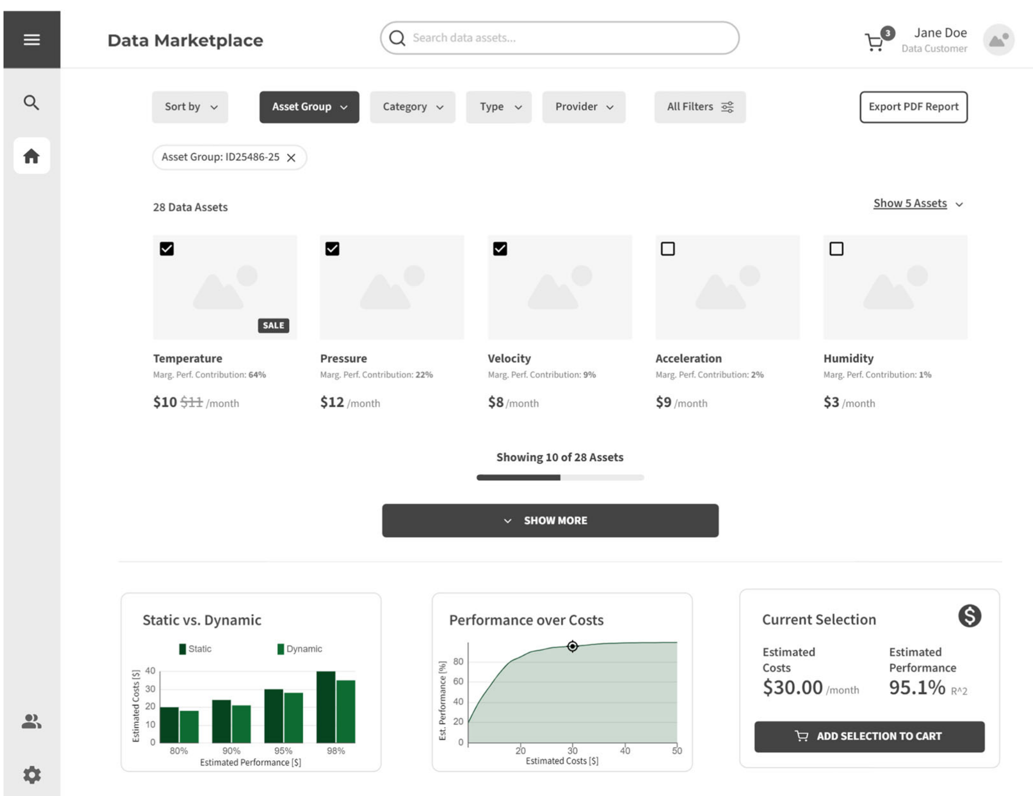
6.1 Towards Implementing and Deploying the DQSM in Data Marketplaces

To illustrate how the technical design knowledge gained in this study would manifest in a novel feature of a data marketplace, we created a mockup prototype (Fig. 14) to demonstrate how the DQSM would benefit data consumers in selecting appropriate data assets, which in turn would reward data providers and marketplace operators through increased usage of the marketplace.

In the search configuration page, the data consumer interacts with the "data asset selection assistant" by providing a historical target dataset of measurements for a variable to be sensed. This can be a continuous parameter (such as temperature) or a discrete characteristic (such as defect or no defect). The user can then choose between data-based or model-based assessment ("level of dependency", see DP2) and, in the case of model-based assessment, upload an existing black-box model or application programming interface access to a deployed model that the



(a) Search configuration page



(b) Result visualization and selection page

Fig. 14 Mockups of DQSM implemented into data marketplace user wants to improve by adding more data. The user must also select whether the value assessment should take into account that data assets are only queried in certain contexts or are fully included ("context-awareness", see DP1). A static evaluation would collectively assess the impact of a particular data asset (e.g., speed) without regard to individual circumstances. In contrast, a dynamic assessment would evaluate the impact of that data asset on a case-by-case basis, considering its particular context (e.g., weather conditions). Finally, the user can use filters to narrow down datasets based on domain knowledge (e.g., only certain car brands supported by the intended service).

Based on these user inputs, the DQSM applies the presented technical implementation modules to quantify the value of different data assets. First, the Data Asset Ranking Module quantifies the predictive power of an available data asset to the provided target in the form of rank scores that are unique per feature in static assessment scenarios and unique per feature value in dynamic selection scenarios. Second, the data asset quantification module applies sequential feature quantification to determine the marginal contribution of a data asset to a given objective function, scaling their contribution to a range between 0 and 1 to translate the importance estimation into actionable insights. Third, the Data Asset Selection Support Module translates these outcomes into concrete outputs for decision-making in the context of data asset acquisition and utilization.

In the second page of the mockup ("result visualization and selection"), these outcomes are then presented to the user to inform his decision-making (DP4), ultimately enhancing the value of the data marketplace for all stakeholders (DP3). The different data assets considered are listed and ranked based on their marginal performance contribution, listing also their cost on the data marketplace. The user can then select the data assets he intends to purchase to view the estimated cost and performance of the current selection. By clicking a button to add the selected data assets to the cart, the user can return to the basic data marketplace view.

Transitioning the DQSM prototype into a scalable, real-world solution within a data marketplace requires a multi-faceted approach that addresses both front-end and back-end considerations. By focusing on user-centric interface design, ensuring seamless marketplace integration, and adopting a modular and scalable architecture, the DQSM can evolve from a research prototype to a real-world solution with the capability to significantly enhance the value proposition of data marketplaces for both data providers and consumers in the future.

First, the front-end requires a user-centric redesign (Lizcano et al. 2009). While the current prototype is functional for research, it must be refined to meet the diverse needs of data marketplace users. Incorporating interactive visualizations and plots can enhance the user experience by allowing them to explore data asset characteristics, relationships, and potential value contributions (Chen et al. 2020). Customizable value metrics that allow users to define weights or incorporate domain-specific value functions can further personalize the experience. The integration of scenario analysis tools, allowing users to simulate the impact of different data asset combinations, would provide valuable insight into potential acquisition

Table 6 Nascent design knowledge for DSQM based on Jones and Gregor (2007)

Component	Description
Purpose and scope	Prescriptive knowledge for developing a Data Asset Value Quantification and Selection Mechanism (DQSM) to alleviate information asymmetry in data marketplaces.
Key Constructs	We defined two core levels of output-specific constructs (Offermann et al. 2010): the level of dependency (model-based vs. data-based) and the context-awareness (static vs. dynamic) – both relating to a set of seven design requirements.
Principles of form and function	Drawing from the body of knowledge, we derived seven design requirements and four tentative design principles and evaluated the design in five evaluation episodes through six design features.
Justificatory knowledge	We conceptualize our design principles based on justificatory knowledge from the field of feature selection techniques, explainable artificial intelligence, and active feature value acquisition.
Testable propositions	We implemented a DQSM that integrates feature selection and XAI techniques is technically feasible and enables accurate and context-sensitive quantification of data asset value, as evidenced by consistent performance across diverse datasets and positive expert feedback on its applicability in real-world scenarios.
Artifact mutability	We discuss the mutability of DQSMs due to recent advances in both machine learning and explainable AI.
Principles of implementation	We suggest implementing the six design features in three different modules as a concrete instantiation of the design principles.
Expository instantiation	We built and evaluated an artifact to support the experts in selecting relevant data assets. Through mockups, we illustrate how the DQSM could be incorporate into a data marketplace application.

strategies before purchase. Figure 14 illustrates a potential user interface that allows marketplace users to interact with the DQSM module. Figure 14a shows how to specify the search and incorporates key elements of the DQSM framework, such as the level of dependency and context-awareness. Figure 14b visualizes the results in an interactive, clear way and provides decision support to the user.

Second, the DQSM front-end should be designed to be understandable to technical non-experts. This involves translating complex technical assessments into understandable terms and clearly communicating the implications of different choices. For example, when choosing among the four suggested scenarios, users should be presented with clear explanations of each option's potential benefits and drawbacks, enabling them to make informed decisions that align with their specific needs and goals. This focus on user comprehension not only improves the usability of the DQSM, but also fosters trust in the data marketplace ecosystem.

Third, back-end integration requires seamless incorporation of the DQSM into the existing data marketplace infrastructure. This involves addressing technical challenges related to scalability, performance, and security – areas that are often already addressed by existing data marketplaces. Optimizing the DQSM algorithm to efficiently handle large volumes of data assets and user requests through techniques such as distributed computing and caching mechanisms is critical. Robust security measures, including encryption, access controls, and privacy compliance, must be implemented to protect the confidentiality and integrity of data assets. In addition, developing or integrating a well-documented API used in the data marketplace would facilitate the integration of the DQSM into various platforms and applications, thereby increasing its reach and potential impact.

Finally, a modular and scalable architecture is recommended for the DQSM. This design, where the DQSM functions as a separate component interacting with the marketplace through a defined API, allows for flexibility and adaptability to different architectures and requirements. The front-end interface communicates with the DQSM via the API, receiving value assessments and visualizations. back-end processes, including data preprocessing, feature extraction, model training, and value calculation, run securely within the DQSM module.

6.2 Contributions to Theory and Practice

This research contributes to both theory and practice in the domain of data marketplaces, and particularly for quantifying the value of data assets and selecting them accordingly. On the theoretical level, it introduces a comprehensive conceptual framework for data asset value quantification and selection within data marketplaces, systematically categorizing and comparing diverse methods derived from feature selection, explainable AI, and active feature-value acquisition. This framework addresses a notable gap in the existing literature by offering a structured approach to understanding the multifaceted nature of data valuation depending on context and model dependency. Moreover, the research explores the integration of ML and XAI methods into the data marketplace context, thereby enhancing transparency and fostering trust in data valuation models. By elucidating the factors that influence value assessments, this integration advances the theoretical understanding of XAI's application and carries practical implications for informed decision-making.

Additionally, the research derives a set of design principles for a DQSM in data marketplaces, addressing theoretically grounded and empirically validated design requirements for such a mechanism. These principles provide actionable guidance for developing mechanisms that address the challenges associated with data asset valuation, selection, and acquisition in data marketplaces, offering valuable insights for platform designers and operators seeking to improve these platforms' functionalities and user experience. Central to this is introducing a novel methodological approach, the DQSM, designed to quantify the value of data assets and guide data marketplace users in selecting suitable data sources.

From a practical standpoint, the proposed DQSM directly addresses the challenges data consumers face in data marketplaces. By equipping consumers with a technically robust and flexible tool for assessing the value and suitability of data assets in their use context, we hope to empower them to make informed purchasing decisions, thereby enhancing the overall efficiency and effectiveness of data marketplaces. The research goes beyond conceptualization – operationalizing data valuation strategies through a technical experiment. This practical demonstration validates the feasibility and potential impact of the proposed approach and serves as a reference point for future research and development efforts in this domain. Importantly, by enabling consumers to assess the potential value of data assets before purchase, the DQSM directly confronts the Arrow Information Paradox (Arrow 1962), a significant obstacle hindering data transactions today (Spiekermann 2019). We hope these practical contributions will support unlocking the full potential of data marketplaces by facilitating more transparent and efficient data exchange.

Furthermore, through an extensive comparison study across various datasets and feature selection strategies, this research offers empirical evidence on the strengths and weaknesses of different technical approaches, guiding practitioners – both data marketplace operators and consumers – in selecting the most suitable method for specific contexts. The integration of diverse ML and XAI techniques demonstrates the versatility and robustness of the proposed artifact for evaluating data assets across different data types and ML models. The study also provides a practical roadmap for incorporating the DQSM into a data marketplace, promoting benefit-oriented pricing strategies, and fostering a balanced and incentivized ecosystem.

The implications of this research also extend beyond data marketplaces, offering valuable insights into the broader field of data monetization and providing guidance for organizations seeking to leverage their data assets effectively. The integration of XAI principles into data valuation processes has the potential to enhance trust and transparency in various data-driven applications, while the design principles for DQSMs can be adapted and applied to other platforms and ecosystems where data sharing plays a crucial role.

6.3 Positioning in the IS Sphere

In their recent ISR editorial, Abbasi et al. (2024) discuss the challenges and opportunities of publishing AI-related research following the DSR paradigm in the IS community. While parts of their guidance are particularly relevant to the application and study of generative AI in research contexts, their editorial also holds valuable insights for conducting and positioning research that applies and studies AI-enabled technologies, such as our research context. The editorial provides guidance on how to address these challenges. It suggests pathways for impactful AI design research, mainly along two core necessities for positioning one's work in the community: the novelty of the solution and its downstream impact.

Regarding novelty, our research introduces an original approach to quantifying data asset value in data marketplaces using, amongst others, ML and XAI methods. This mechanism addresses a critical gap in current marketplace operations (Eckhardt et al. 2022) by providing a transparent and informed means to assess the value of data assets before their acquisition – thus enabling marketplace participants to make better-informed purchasing decisions. In particular, the primary novelty of our research lies in developing the DQSM for potential data marketplaces and applying it to different data assets. By employing ML and XAI methods, we facilitate a deeper understanding of how different data features can contribute to the overall data value. This approach improves the transparency of data valuation and allows for a more nuanced and context-sensitive assessment of data assets. As a consequence, an integration of DQSM helps bridge the gap between data providers and consumers by providing clear insights into the value contribution of individual data assets, which is crucial for informed decision-making in data marketplaces.

From a theoretical perspective, we summarize the contributions of our DQSM artifact in light of the IS design theory framework from Jones and Gregor (2007). Table 6 structures and presents our overall developed nascent design knowledge of DQSM. Additional details are illustrated in Online Appendix A.5.

In terms of downstream impact, our proposed mechanism has significant implications for the functioning of data marketplaces, potentially transforming how data assets are valued and traded. For data providers and marketplace operators, implementing our DQSM can fundamentally alter how data marketplaces display and market data assets. By providing a clear and transparent value assessment of data assets based on the use cases of potential data consumers, our mechanism helps solve the Arrow Information Paradox of asset valuation prior to purchase (Hannila et al. 2022). This practical application can lead to more efficient and effective data sharing, as consumers can better understand and justify the value of their investments. In the long term, our approach can increase participation and engagement in data marketplaces by building trust and reducing uncertainty in data transactions. By ensuring that data providers and consumers clearly understand data value in their specific usage context, our proposed mechanism fosters a more vibrant and dynamic marketplace ecosystem. This increased transparency and trust can lead to greater data sharing and collaboration, ultimately driving innovation and economic growth in data-driven industries.

6.4 Limitations & Future Work

As with any research, this study has inherent limitations. Addressing these limitations offers valuable directions for future research to enhance the reliability and applicability of our framework.

From a design science perspective, measures for demonstrating reliability can hardly ever be complete (Storey et al. 2024), particularly when focusing on designin artifacts such as the DQSM that are part of a larger information system, such as a data marketplace. Our DSR study showed internal and external reliability of our proposed artifact and design knowledge through evaluations based on technical experiments and interviews. However, we acknowledge as a primary limitation of our research that our proposed DQSM has not yet been implemented in a full-scale data marketplace setting. Such an implementation would involve many dependencies and considerations for scalable deployment, as well as a longer-term experimentation with test users to familiarize themselves with the tool. While this would allow for more naturalistic, behavior-based testing of the effectiveness of our artifact and design knowledge in addressing information asymetries and forming trust in the data marketplace, we argue that, given the technology-oriented scope of our research, this would also introduce many contextual factors. These factors could make it harder to isolate the actual effect of the DQSM from the broader data marketplace context.

Therefore, we also acknowledge the limitation that the final evaluation of our artifact relies on user perceptions and attitudes. Although we reported positive feedback, user opinions can be subjective and variable, and may not necessarily translate to actual effectiveness or success in a broader real-world context. We also recognize that some informants involved in the evaluation were part of the goal-setting phase, which may limit external validity to some degree. Future research should include objective measures of effectiveness and expand the evaluation to diverse user groups to validate the findings.

Nonetheless, a primary strength of our DSR study lies in its focus on the technical implementation of the DQSM, evaluated with datasets from multiple domains and yielding promising performance results in the technical experiments. While the datasets used for technical evaluations were selected to be representative, they do not cover all possible data marketplace scenarios. The performance and utility of our framework may differ in scenarios not considered in this study. Future research should test the framework across various scenarios specific to data marketplaces to ensure its robustness and generalizability.

From a technical perspective, a significant limitation is related to the curse of dimensionality and the model-intrinsic feature selection mechanisms, where an asset is considered only if deemed relevant by the model. This necessitates an initial data-based pre-selection to cleanse the selection of less useful data assets. Future research could explore more sophisticated pre-selection algorithms or develop novel feature selection methods to enhance the efficacy of this step.

Additionally, while we aimed to create a generalizable framework, the specific models chosen for the analysis could impact the results. We utilized a variety of models, ranging from simple to complex, without any hyperparameter tuning. More specialized models might yield different insights, and the performance of our framework could vary accordingly. Hence, future studies should investigate the effects of hyperparameter tuning and explore advanced models to assess their impact on the framework's performance.

Another aspect involves the computational intensity of some components, particularly those involving local XAI methods. These can pose challenges in large-scale, real-world applications. This issue could be mitigated by adopting more efficient sampling strategy, as suggested by Strumbelj and Kononenko (2010). Therefore, future research should focus on developing and testing these strategies to enhance computational efficiency.

While this work proposes the utilization of XAI techniques and assumes they work as intended, it is important to note that XAI – especially post-hoc surrogate models – are imperfect (Morrison et al. 2024). Explanations can mislead (Lakkaraju and Bastani 2020), affect human perception and behavior both positive and negative ways (Schoeffer et al. 2024), and have an ambivalent relationship to fairness considerations (Deck et al. 2024).

A further limitation pertains to the assumption of available target labels. In model-based methods, we assume that at least a certain number of observations of the desired target variable are available, focusing exclusively on supervised methods. Exploring unsupervised or semi-supervised methods (Morichetta et al. 2019; Montavon et al. 2020; Wickramasinghe et al. 2021) could broaden the framework's applicability to scenarios with limited or no target labels, which is beneficial for future studies beyond the current scope.

Moreover, relying on strict value-based pricing without considering the actual costs of a data asset poses another limitation of the current prototype. Data assets that do not significantly contribute to a consumer-specific objective might be supplied at a marginal price, risking exploitation by platform users. Developing strategies to delineate non-relevant data assets and pricing them accordingly is crucial for real-world implementation. Future research could focus on creating and testing pricing models that account for the varying significance of data assets.

These limitations present opportunities for future research to improve the reliance and applicability of our framework and further validate its effectiveness across a broader range of real-world data marketplace scenarios. Addressing these limitations will help refine the framework and ensure its success in diverse applications.

7 Conclusion

Data assets are playing an increasingly recognized as critical resources for organizations. However, the potential for exchanging and monetizing such data to drive innovation remains underutilized due to insufficient incentives for providers and limited offerings for consumers, among other reasons. This creates a vicious circle that impedes the adoption of data marketplaces and similar digital platforms.

In response, we propose a novel mechanism for data marketplaces that (1) supports consumers in identifying relevant data assets and making informed buying decisions, (2) thereby generates more demand that incentivizes providers to provide more data assets, and (3) aligns the supply and demand of data assets. Our DQSM represents a decision support tool for selecting data assets during the innovation and implementation of data-driven services, such as the development of virtual sensors. Virtual sensors synthesize higher-order observations of reality from diverse physical sensor signals, but selecting appropriate input data poses a significant challenge due to the abundance of potentially relevant data assets.

Our mechanism addresses this challenge facilitating the selection of relevant data assets, thus supporting the innovation and implementation of data-driven services such as virtual sensors. Guided by DSR, we design a DQSM artifact rooted in justificatory knowledge from the field of ML and XAI. We instantiate and evaluate our artifact employing a technical experiment using publicly available and industry data, comparing various scenarios and methods to identify effective strategies for selecting and quantifying data assets. As a result, we derive and validate four central DPs for a DQSM as generalizable design-oriented knowledge.

This study, demonstrates that quantifying and selecting data assets – essential for innovating and implementing data-driven services – can be encapsulated within a decision-support mechanism. Our proposed mechanism can be integrated into a data marketplace as a matchmaking and pricing feature, promoting benefit-oriented data asset pricing. This approach not only incentivizes both consumers and providers but also fosters increased adoption of these platforms.

Funding Open Access funding enabled and organized by Projekt DEAL.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12599-025-00940-8.

References

Abbas AE, Agahari W, Van de Ven M, Zuiderwijk A, De Reuver M (2021) Business data sharing through data marketplaces: a systematic literature review. J Theor Appl Electron Commer Res 16(7):3321–3339

Abbasi A, Parsons J, Pant G, Sheng ORL, Sarker S (2024) Pathways for design research on artificial intelligence. Inf Syst Res 35(2):441–459

Abraham R, Schneider J, vom Brocke J (2023) A taxonomy of data governance decision domains in data marketplaces. Electron Market 33(1):22

Ackoff R (1989) From data to wisdom. J Appl Syst Anal 16(1):3–9

Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52

Aeberhard S, Forina M (1991) Wine. UCI machine learning repository. https://doi.org/10.24432/C5PC7J

Agarwal A, Dahleh M, Sarkar T (2019) A marketplace for data. In: Proceedings of the 2019 acm conference on economics and computation, New York, NY, USA, pp 701–726

Aha DW, Bankert RL (1995) A comparative evaluation of sequential feature selection algorithms. In: Pre-proceedings of the fifth international workshop on artificial intelligence and statistics, PMLR, pp 1–7

Altmann A, Tolosi L, Sander O, Lengauer T (2010) Permutation importance: a corrected feature importance measure. Bioinform 26(10):1340–1347

Anthony B Jr (2023) Decentralized brokered enabled ecosystem for data marketplace in smart cities towards a data sharing economy. Environ Syst Decis 43(3):453–471

Apley DW, Zhu J (2020) Visualizing the effects of predictor variables in black box supervised learning models. J Royal Stat Soc Series B Stat Methodol 82(4):1059–1086

Arrieta AB, Rodriguez ND, Ser JD, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115

Arrow K (1962) Economic welfare and the allocation of resources for invention. The rate and direction of inventive activity: economic and social factors. Princeton University Press, Princeton, pp 609–626

Azcoitia SA, Laoutaris N (2022) A survey of data marketplaces and their business models. ACM SIGMOD Record 51(3):18–29

Badewitz W, Hengesbach C, Weinhardt C (2022) Challenges of pricing data assets: a literature review. In: 2022 IEEE 24th conference on Business Informatics (CBI), IEEE, vol 1, pp 80–89

Baehrens D, Schroeter T, Harmeling S, Kawanabe M, Hansen K, Müller K (2010) How to explain individual classification decisions. J Mach Learn Res 11:1803–1831

Bastiaansen H, Dalmolen S, Kollenstart M, Punter M (2019) Infrastructural sovereignty over agreement and transaction data ('metadata') in an open network-model for multilateral sharing of sensitive data. In: 40th international conference on information systems, ICIS 2019

Becker B, Kohavi R (1996) Adult. UCI machine learning repository. https://doi.org/10.24432/C5XW20

Bellman R (1966) Dynamic programming. Science 153(3731):34–37

Bermingham ML, Pong-Wong R, Spiliopoulou A, Hayward C, Rudan I, Campbell H, Wright AF, Wilson JF, Agakov F, Navarro P et al (2015) Application of high-dimensional feature selection: evaluation for genomic prediction in man. Sci Rep 5(1):1–12

Beverungen D, Kundisch D, Wünderlich N (2021) Transforming into a platform provider: strategic options for industrial smart service providers. J Serv Manag 32(4):507–532

Bissyandé TF, Thung F, Lo D, Jiang L, Réveillère L (2013) Popularity, interoperability, and impact of programming languages in 100, 000 open source projects. In: 37th annual IEEE computer software and applications conference, COMPSAC 2013. Kyoto, Japan, pp 303–312

Blum A, Langley P (1997) Selection of relevant features and examples in machine learning. Artif Intell 97(1–2):245–271

Bommert A, Sun X, Bischl B, Rahnenführer J, Lang M (2020) Benchmark for filter methods for feature selection in high-dimensional classification data. Comput Stat Data Anal 143:106839

Breiman L (2001) Statistical modeling: the two cultures. Technical report

Brown G, Pocock AC, Zhao M, Luján M (2012) Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. J Mach Learn Res 13:27–66

Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, Niculae V, Prettenhofer P, Gramfort A, Grobler J, Layton R, Vanderplas J, Joly A, Holt B, Varoquaux G (2013) API design for machine learning software: experiences from the scikit-learn project. arXiv:1309.0238

Campana MG, Chatzopoulos D, Delmastro F, Hui P (2018) Lightweight modeling of user context combining physical and virtual sensor data. In: Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers, UbiComp/ISWC 2018 Adjunct, Singapore, pp 1309–1320

Chen L, Koutris P, Kumar A (2019) Towards model-based pricing for machine learning in a data marketplace. In: Proceedings of the international conference on management of data, SIGMOD conference, Amsterdam, the Netherlands, pp 1535–1552

Chen M, Hauser H, Rheingans P, Scheuermann G (2020) Foundations of data visualization. Springer, Heidelberg

Chen T, Guestrin, C (2016) Xgboost: a scalable tree boosting system. In: Proceedings of the 22Nd ACM SIGKDD international conference on knowledge discovery and data mining (New York, NY, USA, 2016), KDD 16, ACM, pp 785–794

Choi SKT, Kröschel I (2015) Challenges of governing interorganizational value chains: insights from a case study. In: 23rd European conference on information systems, ECIS 2015

Chollet F (2015) Keras. https://github.com/fchollet/keras. Accessed 5 Nov 2023

Chowdhury MJM, Kayes A, Watters PA, Scolyer-Gray P, Ng A, Dillon T (2020) Patient controlled, privacy preserving IoT healthcare data sharing framework. In: Proceedings of the Hawaii international conference on system sciences (HICSS), pp 1–11

Constantinides P, Henfridsson O, Parker GG (2018) Introduction – platforms and infrastructures in the digital age. Inf Syst Res 29(2):381–400

Craven MW, Shavlik JW (1996) Extracting thee-structured representations of thained networks. Adv Neural Inf Process Syst pp 24–30

Cusumano MA, Yoffie DB, Gawer A (2020) The future of platforms. MIT Sloan Manag Rev pp 26–34

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q pp 319–340

Deck L, Schoeffer J, De-Arteaga M, Kühl N (2024) A critical survey on fairness benefits of explainable AI. In: Proceedings of the ACM conference on fairness, accountability and transparency (FAccT)

Dorogush AV, Ershov V, Gulin A (2018) CatBoost: gradient boosting with categorical features support. arXiv:1810.11363

Doshi-Velez F, Kim B (2017) Towards a rigorous science of interpretable machine learning. arXiv:1702.08608

Eckhardt S, Sprenkamp K, Zavolokina L, Bauer I, Schwabe G (2022) Can artificial intelligence help used-car dealers survive in a data-driven used-car market? International conference on design science research in information systems and technology. Springer, Heidelberg, pp 115–127

Enders T, Martin D, Sehgal GG, Schüritz R (2020) Igniting the spark: overcoming organizational change resistance to advance innovation adoption – the case of data-driven services. In: Exploring service science – proceedings of the 10th international conference, IESS 2020, Porto, Portugal, Lecture notes in business information processing, vol 377, pp 217–230

Faruqui A, Harris D, Hledik R (2010) Unlocking the 53 billion savings from smart meters in the EU: how increasing the adoption of dynamic tariffs could make or break the EU's smart grid investment. Energy Policy 38(10):6222–6231

Fassnacht M, Benz C, Heinz D, Leimstoll J, Satzger G (2023) Barriers to data sharing among private sector organizations. In: 56th Hawaii international conference on system sciences HICSS. Maui, Hawaii, USA, pp 3695–3704

Fassnacht M, Leimstoll J, Benz C, Heinz D, Satzger G (2024) Data sharing practices: the interplay of data, organizational structures, and network dynamics. Electron Market 34(1):1–25

Fayyad UM, Piatetsky-Shapiro G, Smyth P (1996) The KDD process for extracting useful knowledge from volumes of data. Commun ACM 39(11):27–34

Ferri FJ, Pudil P, Hatef M, Kittler J (1994) Comparative study of techniques for large-scale feature selection. In: Machine intelligence and pattern recognition, vol 16, Elsevier, pp 403–413

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232

Fruhwirth M, Rachinger M, Prlja E (2020) Discovering business models of data marketplaces. In: 53rd Hawaii international conference on system sciences (HICSS)

Gama J (2010) Knowledge discovery from data streams. CRC Press, Chapman and Hall / CRC Data Mining and Knowledge Discovery Series

Gandomi A, Haider M (2015) Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manag 35(2):137–144

Gasco-Hernandez M, Feng W, Gil-Garcia JR (2018) Providing public value through data sharing: understanding critical factors of food traceability for local farms and institutional buyers. In: 51st Hawaii international conference on system sciences (HICSS) 2018

Gelhaar J, Otto B (2020) Challenges in the emergence of data ecosystems. In: Pacific Asia Conference on Information Systems (PACIS)

Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter MA, Kagal L (2018) Explaining explanations: an overview of interpretability of machine learning. In: 5th IEEE international conference on data science and advanced analytics, DSAA 2018. Italy, Turin, pp 80–89

Grover V, Chiang RHL, Liang TP, Zhang D (2018) Creating strategic business value from big data analytics: a research framework. J Manag Inf Syst 35(2):388–423. https://doi.org/10.1080/07421222.2018.1451951

Guggenberger TM, Hunke F, Möller F, Eimer AC, Satzger G, Otto B (2021) How to design IIoT-platforms your partners are eager to join: learnings from an emerging ecosystem. In: Proceedings of the 16th international conference on wirtschafts informatik (WI)

Günther LC, Colangelo E, Wiendahl HH, Bauer C (2019) Data quality assessment for improved decision-making: a methodology for small and medium-sized enterprises. In: Procedia manufacturing 29:583–591

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Lear Res 3:1157–1182

Hagiu A, Wright J (2015) Multi-sided platforms. Int J Ind Organ 43:162–174

Hall MA (2000) Correlation-based feature selection for discrete and numeric class machine learning. In: Proceedings of the seventeenth international conference on machine learning (ICML 2000), Stanford University, Stanford, CA, USA, Morgan Kaufmann, pp 359–366

Hancock JT, Khoshgoftaar TM (2020) Survey on categorical data for neural networks. J Big Data 7(1):1–41

Hannila H, Silvola R, Harkonen J, Haapasalo H (2022) Data-driven begins with data; potential of data assets. J Comput Inf Syst 62(1):29–38

Harrison D, Rubinfeld D (1978) Boston housing. StatLib repository. http://lib.stat.cmu.edu/datasets/boston. Accessed 05 Nov 2023

Hevner AR (2007) A three cycle view of design science research a three cycle view of design science research. Scand J Inf Syst 19(2):1–6

Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. MIS Q 28(1):75–105

Hirt R, Kühl N, Martin D, Satzger G (2023) Enabling inter-organizational analytics in business networks through meta machine learning. Information technology management, pp 1–25

Huang L, Dou Y, Liu Y, Wang J, Chen G, Zhang X, Wang R (2021) Toward a research framework to conceptualize data as a factor of production: the data marketplace perspective. Fundam Res 1(5):586–594

Hunke F, Heinz D, Satzger G (2022) Creating customer value from data: foundations and archetypes of analytics-based services. Electron Market 32(2):503–521

Hynes N, Dao D, Yan D, Cheng R, Song D (2018) A demonstration of sterling: a privacy-preserving data marketplace. In: Proceedings of the VLDB endowment, 11(12):2086–2089

Ilyas EB, Fischer M, Iggena T, Tönjes R (2020) Virtual sensor creation to replace faulty sensors using automated machine learning techniques. 2020 global internet of things summit. GIoTS, Dublin, Ireland, pp 1–6

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning, vol 112. Springer

Janisch J, Pevný T, Lisý V (2019) Classification with costly features using deep reinforcement learning. Thirty-third AAAI Conference on Artificial Intelligence, AAAI 2019, the Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, the Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019. Honolulu, Hawaii, USA, pp 3959–3966

Janosi A, Steinbrunn W, Pfisterer M, Detrano R (1988) Heart disease. UCI machine learning repository. https://doi.org/10.24432/C52P4X

Jones D, Gregor S (2007) The anatomy of a design theory. J Assoc Inf Syst 8(5):19

Jussen I, Fassnacht M, Schweihoff JC, Möller F (2024) Reaching for the stars: exploring value constellations in inter-organizational data sharing. In: 32nd European conference on information systems (ECIS)

Kabadayi S, Pridgen A, Julien C (2006) Virtual sensors: abstracting data from physical sensors. In: International symposium on a world of wireless, mobile and multimedia networks (WoWMoM), Buffalo, New York, USA, proceedings, pp 587–592

Kachuee M, Goldstein O, Kärkkäinen K, Darabi S, Sarrafzadeh M (2019) Opportunistic learning: budgeted cost-sensitive learning from data streams. In: 7th international conference on learning representations, ICLR, New Orleans, LA, USA

Kahn M (1994) Diabetes. UCI machine learning repository. https://doi.org/10.24432/C5T59G

Kajüter P, Arlinghaus T, Kus K, Teuteberg F (2022) Analysis of barriers to digital linking among healthcare stakeholders. In: Wirtschaftsinformatik 2022 proceedings

Kanani P, Melville P (2008) Prediction-time active feature-value acquisition for cost-effective customer targeting. Advances in neural information processing systems (NIPS)

Kawakami T, Ly BLN, Takeuchi S, Teranishi Y, Harumoto K, Nishio S (2008) Distributed sensor information management architecture based on semantic analysis of sensing data. In: Proceedings of the international symposium on applications and the internet, SAINT, Turku, Finland, pp 353–356

Kaya H, Tüfekci P, Uzun E (2019) Gas turbine CO and NOx emission data set. UCI machine learning repository. https://doi.org/10.24432/C5WC95

Klaise J, Van Looveren A, Vacanti G, Coca A (2021) Alibi: algorithms for monitoring and explaining machine learning models. https://github.com/SeldonIO/alibi. Accessed 05 Nov. 2023

Klein A, Lehner W (2009) Representing data quality in sensor data streaming environments. J Data Infn Qual 1(2):1–28

Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97(1–2):273–324

Koutroumpis P, Leiponen A, Thomas L (2017) The (unfulfilled) potential of data marketplaces. http://pub.etla.fi/ETLA-Working-Papers-53.pdf. Accessed 05 Nov. 2023

Kozodoi N, Lessmann S, Papakonstantinou K, Gatsoulis Y, Baesens B (2019) A multi-objective approach for profit-driven feature selection in credit scoring. Decis Support Syst, 120:106–117

Kuechler B, Vaishnavi V (2008) On theory development in design science research: anatomy of a research project. Eur J Inf Syst 17(5):489–504

Kühl N (2024) Human-centric artificial intelligence: the road ahead. Transf Zeitschr Kommun Markenmanag, 70(4)

Kühl N, Hirt R, Baier L, Schmitz B, Satzger G (2021) How to conduct rigorous supervised machine learning in information systems research: the supervised machine learning report card. Commun Assoc Inf Syst 48:46

Kühl N, Schemmer M, Goutier M, Satzger G (2022) Artificial intelligence and machine learning. Electron Market 32(4):2235–2244

Kumar V, Minz S (2014) Feature selection: a literature review. Smart Comput Rev 4(3):211–229

Lakkaraju H, Bastani O (2020) "How do i fool you?" Manipulating user trust via misleading black box explanations. In: Proceedings of the AAAI/ACM conference on AI, ethics, and society, pp 79–85

LeCun Y, Bengio Y, Hinton GE (2015) Deep learning. Nature 521(7553):436–444

Lee C, Birch D, Wu C, Silva D, Tsinalis O, Li Y, Yan S, Ghanem M, Guo Y (2013) Building a generic platform for big sensor data application. In: 2013 IEEE international conference on big data. Santa Clara, CA, USA, pp 94–102

Lee J, Bagheri B, Jin C (2016) Introduction to cyber manufacturing. Manufact Lett 8:11–15

Lessmann S, Voß S (2009) Feature selection in marketing applications. In: Advanced data mining and applications: 5th international conference, ADMA 2009, Beijing, China, proceedings 5. Springer, Heidelberg, pp 200–208

Li C, Wang X, Dong W, Yan J, Liu Q, Zha H (2019) Joint active learning with feature selection via CUR matrix decomposition. IEEE Transact Pattern Anal Mach Intell 41(6):1382–1396

Liu C, White RW, Dumais ST (2010) Understanding web browsing behaviors through weibull analysis of dwell time. In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval, Geneva, Switzerland, pp 379–386

Lizcano D, Soriano J, Reyes M, Hierro JJ (2009) A user-centric approach for developing and deploying service front-ends in the future internet of services. Int J Web Grid Serv 5(2):155–191
Lundberg SM, Lee S (2017) A unified approach to interpreting model predictions. In: Advances in neural information processing systems 30: annual conference on neural information processing systems. Long Beach, CA, USA, pp 4765–4774
Ma C, Tschiatschek S, Palla K, Hernández-Lobato JM, Nowozin S, Zhang C (2019) EDDI: efficient dynamic discovery of high-value information with partial VAE. In: Proceedings of the 36th international conference on machine learning, ICML, Long Beach, California, USA, proceedings of machine learning research, vol 97, pp 4234–4243
Madria S, Kumar V, Dalvi R (2014) Sensor cloud: a cloud of virtual sensors. IEEE Softw 31(2):70–77
March ST, Smith GF (1995) Design and natural science research on information technology. Decision Support Syst 15(4):251–266
Martin D, Kühl N, Satzger G (2021) Virtual sensors. Bus Inf Syst Eng 63(3):315–323
McAfee A, Brynjolfsson E, Davenport TH, Patil D, Barton D (2012) Big data: the management revolution. Harvard Bus Rev 90(10):60–68
Melville N, Kraemer K, Gurbaxani V (2004) Review: information technology and organizational performance: an integrative model of it business value. MIS Q 28(2):283–32. https://doi.org/10.2307/25148636
Melville P, Provost FJ, Mooney RJ (2005) An expected utility approach to active feature-value acquisition. In: Proceedings of the 5th IEEE international conference on data mining (ICDM), Houston, Texas, USA, pp 745–748
Miller HG, Mork P (2013) From data to decisions: a value chain for big data. IT Prof 15(1):57–59
Misura K, Zagar M (2016) Data marketplace for internet of things. In: International conference on smart systems and technologies (SST), pp 255–260
Molnar C (2019) Interpretable machine learning: a guide for making black box models explainable
Montavon G, Kauffmann JR, Samek W, Müller K (2020) Explaining the predictions of unsupervised learning models. In: xxAI – beyond explainable AI – international workshop, held in Conjunction with ICML, Vienna, Austria, revised and extended papers, LNCS, vol 13200, pp 117–138
Morichetta A, Casas P, Mellia M (2019) EXPLAIN-IT: towards explainable ai for unsupervised network traffic analysis. In: Proceedings of the 3rd ACM CoNEXT workshop on big data, machine learning and artificial intelligence for data communication networks, Big-DAMA at CoNEXT,Orlando, FL, USA, pp 22–28
Morrison K, Spitzer P, Turri V, Feng M, Kühl N, Perer A (2024) The impact of imperfect XAI on human-AI decision-making. In: Proceedings of the ACM on human-computer interaction, 8(CSCW1):1–39
Offermann P, Blom S, Levina O, Bub U (2010) Proposal for components of method design theories: increasing the utility of method design artefacts. Bus Inf Syst Eng 2:295–304
Pace K, Barry R (1997) California housing. StatLib repository. https://www.dcc.fc.up.pt/~ltorgo/Regression/cal%5Fhousing.html. Accessed 05 Nov. 2023
Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K (2015) Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Admin Policy Mental Health Res 42(5):533–544
Pant V, Yu E (2018) Getting to win-win in industrial collaboration under coopetition: a strategic modeling approach. Perspectives in business informatics research: 17th international conference, BIR, Stockholm, Sweden, proceedings 17. Springer, Heidelberg, pp 47–66
Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2008) A design science research methodology for information systems research. J Manag Inf Syst 24(3):45–77
Peffers K, Rothenberger M, Tuunanen T, Vaezi R (2012) Design science research evaluation. Design science research in information systems advances in theory and practice, pp 398–410
Petter S, Khazanchi D, Murphy JD (2010) A design science based evaluation framework for patterns. ACM SIGMIS Database: DATABASE Adv Inf Syst 41(3):9–26
Pries-Heje J, Baskerville RL, Venable JR (2008) Strategies for design science research evaluation. In: 16th European conference on information systems. ECIS, Galway, Ireland, pp 255–266
Pudil P, Novovicová J, Kittler J (1994) Floating search methods in feature selection. Pattern Recogn Lett 15(10):1119–1125
Quinlan R (1993) Auto MPG. UCI machine learning repository. https://doi.org/10.24432/C5859H
Ribeiro MT, Singh S, Guestrin C (2016a) Model-agnostic interpretability of machine learning. arXiv:1606.05386
Ribeiro MT, Singh S, Guestrin C (2016b) "Why should I trust you?": explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, San Francisco, CA, USA, pp 1135–1144
Rowley J (2007) The wisdom hierarchy: representations of the DIKW hierarchy. J Inf Sci 33(2):163–180
Saar-Tsechansky M, Provost F (2009) Active feature-value acquisition active feature-value acquisition. Manag Sci 55:664–684
Saeys Y, Inza I, Larranaga P (2007) A review of feature selection techniques in bioinformatics. Bioinform 23(19):2507–2517
Sahakyan M, Aung Z, Rahwan T (2021) Explainable artificial intelligence for tabular data: a survey. IEEE Access 9:135
Salminen M (2018) A metadata model for hybrid data products on a multilateral data marketplace. PhD Thesis
Samek W, Wiegand T, Müller K (2017) Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv:1708.08296
Samuelson PA (1954) The pure theory of public expenditure. Rev Econ Stat, pp 387–389
Schoeffer J, De-Arteaga M, Kuehl N (2024) Explanations, fairness, and appropriate reliance in human-AI decision-making. In: Proceedings of the CHI conference on human factors in computing systems, pp 1–18
Schomm F, Stahl F, Vossen G (2013) Marketplaces for data: an initial survey. ACM SIGMOD Record 42(1):15–26
Schryen G (2013) Revisiting IS business value research: what we already know, what we still need to know, and how we can get there. Eur J Inf Syst 22(2):139. https://doi.org/10.1057/ejis.2012.45
Shapley LS (1953) A value for N-person games. Contrib Theory Games 2(28):307–317
Sober M, Scaffino G, Schulte S, Kanhere SS (2022) A blockchain-based IoT data marketplace. Cluster Comput, pp 1–23
Spiekermann M (2019) Data marketplaces: trends and monetisation of data goods. Intereconomics 54(4):208–216
Stahl F, Schomm F, Vossen G, Vomfell L (2016) A classification framework for data marketplaces. Vietnam J Comput Sci 3(3):137–143
Sterk F, Heinz D, Hengstler P, Weinhardt C (2023) Reallocating uncertainty in incumbent firms through digital platforms: the case of google's automotive ecosystem involvement. In: Proceedings of the International conference on information systems (ICIS)
Storey VC, Baskerville RL, Kaul M (2024) Reliability in design science research. Inf Syst J

Strumbelj E, Kononenko I (2010) An efficient explanation of individual classifications using game theory. J Mach Learn Res 11:1–18

Susha I, Gil-Garcia JR (2019) A collaborative governance approach to partnerships addressing public problems with private data. In: 52nd Hawaii international conference on system sciences (HICSS) 2019

Thorndike RL (1953) Who belongs in the family? Psychometrika 18(4):267–276

Van Alstyne MW, Parker GG, Choudary SP (2016) Pipelines, platforms, and the new rules of strategy. Harvard Bus Rev 94(4):54–62

Vargo SL, Lusch RF (2008) Service-dominant logic: continuing the evolution. J Acad Mark Sci 36:1–10

Vargo SL, Lusch RF (2018) The SAGE handbook of service-dominant logic. Sage, Thousand Oaks

Venable J, Pries-Heje J, Baskerville R (2016) FEDS: a framework for evaluation in design science research. Eur J Inf Syst 25(1):77–89

Vesselkov A, Hämmäinen H, Töyli J (2019) Design and governance of mhealth data sharing. Commun Assoc Inf Syst 45:299–321

Wang F, Liu J (2011) Networked wireless sensor data collection: issues, challenges, and approaches. IEEE Commun Survey Tutor 13(4):673–687

Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. MIS Q 26(2)

Wickramasinghe CS, Amarasinghe K, Marino DL, Rieger C, Manic M (2021) Explainable unsupervised machine learning for cyber-physical systems. IEEE Access 9:131

Wixom BH, Ross JW (2017) How to monetize your data. MIT Sloan Manag Rev 58(3)

Wolberg W, Mangasarian O, Street N, Street W (1995) Breast cancer Wisconsin (diagnostic). UCI machine learning repository. https://doi.org/10.24432/C5DW2B

Yang C, Rangarajan A, Ranka S (2018) Global model interpretation via recursive partitioning. In: 20th IEEE international conference on high performance computing and communications; 16th IEEE international conference on smart city; 4th IEEE international conference on data science and systems. HPCC/SmartCity/DSS, Exeter, United Kingdom, pp 1563–1570

Zeiringer J (2021) Tackling knowledge risks in data-centric collaborations: a literature review. In: 25th Pacific Asia conference on information systems (PACIS) 2021, p 37

Zhang M, Beltrán F, Liu J (2023) A survey of data pricing for data marketplaces. IEEE Transact Big Data 9(4):1038–1056

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J Royal Stat Soc Series B Stat Methodol 67(2):301–320

Zuiderwijk A, Loukis E, Alexopoulos C, Janssen M, Jeffery K (2014) Elements for the development of an open data marketplace. In: Conference for E-democracy and open government, pp 309–322