Proceedings of the 57th Hawaii International Conference on System Sciences | 2024

Supporting Managerial Decision-Making for Federated Machine Learning: Design of a Technology Selection Tool

Milena Zahn
Technical University of Munich
and SAP SE
milena.zahn@sap.com

Tobias Müller
Technical University of Munich
and SAP SE
tobias.mueller15@sap.com

Florian Matthes
Technical University of Munich
matthes@tum.de

Abstract

The insufficient amount of training data is a persisting bottleneck of Machine Learning systems. A large portion of the world's data is scattered and locked in data silos. Breaking up these data silos could alleviate this problem. Federated Machine Learning is a novel model-to-data approach that enables the training of Machine Learning models, on decentralized, potentially siloed data. Despite its promising potential, most Federated Machine Learning projects never leave the prototype stage. This can be attributed to exaggerated expectations and an inappropriate fit between the technology and the use case. Current literature does not offer guidance for assessing the fit between Federated Machine Learning and their use case. Against this backdrop, we design a decision-support tool to aid decision-makers in the suitability and complexity assessment of FedML projects. Thereby, we aim to facilitate the technology selection process, avoid exaggerated expectations and consequently facilitate the success of Federated Machine Learning projects.

Keywords: Federated Machine Learning, Technology Adoption, Design Science Research

1. Introduction

The lack of sufficient and high-quality training data is a persisting challenge in engineering production-ready Machine Learning (ML) systems. Especially small and medium-sized enterprises (SMEs) suffer from an insufficient amount of training data for the development of data-demanding ML systems (Bauer et al., 2020). Despite the growing wealth of digitized data, a considerable amount is still unavailable, scattered, and locked up in data silos. Especially SMEs could facilitate the lack of training data by breaking down these data silos through sharing data and collaborating. However, the companies' willingness to share data is low due to privacy concerns and potential loss of intellectual property (Schomakers et al., 2020).

FedML is a novel ML paradigm that allows the joint training of an ML model on distributed data without the direct need for data sharing. Through its model-to-data approach, FedML allows organizations to collaborate on ML projects without having to disclose their data to other organizations. As pointed out by the World Economic Forum (2020), this capability to collaborate and share data will become increasingly important as ”true masters of digitalization” not only leverage their own data but also improve existing applications or create new ones with data collaboration.

Despite its promise to foster collaboration and enable the usage of currently untapped data, the adoption of FedML in production-ready systems remains limited (Lo et al., 2021). The lack of operationalized FedML systems can be attributed to a manifold of factors, such as the technical complexity of engineering non-deterministic ML systems (Giray, 2021) or the difficulties of managing collaborative projects (Müller et al., 2023). Additionally, the complexity and number of emerging technologies make it increasingly complicated for practitioners and decision-makers to get a solid understanding of the technology that is needed to determine the appropriate fit of a technology for their use case (Shen et al., 2010). This contributes to the observation of Maghazei et al. (2022), that decision-makers tend to use emerging technologies only based on the hype surrounding the technology without examining their actual business benefits. However, a well-grounded fit between the

URI: https://hdl.handle.net/10125/107193
978-0-9981331-7-1

use case and the technology is the fundamental basis to avoid exaggerated expectations (Alsheibani et al., 2018), generate value, and in consequence facilitate the success of the project. Therefore, it gets increasingly important to support the understanding of complex technologies and provide managers with decision-support tools that aid in making a well-grounded technology selection decision.

The current literature corpus does not offer guidance or decision-support tools for practitioners to assess the suitability and complexity of FedML use cases. This research gap presents a significant challenge for decision-makers and practitioners in the successful and efficient implementation of collaborative Artificial Intelligence (AI) projects. We aim to close this research gap by exploring how practitioners can be supported in the technology selection process of FedML projects and by designing a corresponding decision-support tool.

This work intends to aid practitioners in making grounded decisions for using FedML in their use case. Thereby, we intend to help in the successful project implementation and creation of substantive value through FedML. Summarized, we aim to answer the following research questions (RQs):

RQ1: How can practitioners be supported in the technology selection process of FedML projects?

RQ2: How can a corresponding decision-support tool be designed for assessing the feasibility and complexity of FedML projects?

Following, we present theoretical background and related work on FedML as well as technology selection in Section 2. As described in Section 3, we aim to answer the RQs by following Design Science Research (DSR). Section 4 describes the results structured according to the DSR steps. Finally, we discuss the contributions and limitations in Section 5 and conclude with an outline of future research in Section 6.

2. Theoretical Background and Related Work

The following section presents the theoretical background of our study and presents related work. We first describe preliminaries on FedML, followed by the motivation, background, and related work on technology selection processes and tools.

2.1. Federated Machine Learning

FedML is a novel ML technique that enables the collaborative training of a joint ML model on distributed datasets without the need of sharing data. In traditional ML settings, the data is usually collected in a central location, where the ML model is subsequently trained. Hence, data owners need to share their data with a central server and thereby risk losing their intellectual property. FedML counteracts this need of sharing datasets through a model-to-data approach.

Introduced by McMahan et al. (2016), the basic FedML algorithm can be divided into four distinct steps. These steps are illustrated in Figure 1. In the first step, it is required to select an ML model architecture that is suitable for the use case and underlying data structure. Optionally, this initial global model can be pre-trained on a suitable dataset before the model is federated. Secondly, the global model is distributed amongst all participating clients. Thirdly, each client trains the global model on its own local dataset and stores the updated model parameters. Thereby, each client produced an individual version of the global model based on the clients' local dataset. Lastly, each client sends their updated model parameters back to the server for aggregation. The server collects and combines the received model parameters through a pre-defined aggregation protocol such as weighted averaging. The aggregated result of the locally computed parameters is then used to update the global model. These steps can be repeated until a certain accuracy level is reached or until the accuracy converges.

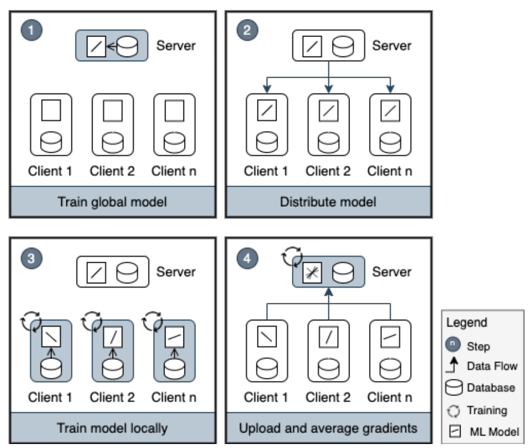

Figure 1. Federated Machine Learning process.

2.2. Technology Selection

Technology selection aims to identify the most suitable technology for a specific need. As technologies get more complex, technology selection becomes a more cumbersome and less predictable course. It gets increasingly difficult to determine the appropriate fit of a technology for a given use case (Shen et al., 2010). Additionally, emerging technologies such as FedML can be described as general-purpose technologies. This means that ML models have various application opportunities and accordingly, possible use cases are not always directly obvious. Therefore, a more opportunistic perspective can enable the creation of entirely new use cases (Jöhnk et al., 2021). In contrast to previous approaches, companies are now also challenged to additionally identify reliable use cases to stay competitive (Hofmann et al., 2020). Hence, analyzing the interplay between *technology push* from a thriving ecosystem, and *market pull* from companies seeking business value becomes increasingly important (Maghazei et al., 2022; Vorraber et al., 2019). A solid understanding of the technology is important before identifying use cases to avoid exaggerated expectations and effective usage (Alsheibani et al., 2018).

Current literature provides several decision-support artifacts to aid practitioners in evaluating the suitability of a technology for their particular use case. On the example of Blockchain, Wüst and Gervais (2018) proposed a decision tree that critically analyzed whether blockchain is a reasonable technology for the underlying use case. The authors argue, that the hype and promise of blockchain lead to overuse and therefore focus on when a use case does not need the adoption of the technology. Similar studies provided decision support on whether blockchain may be the appropriate technical infrastructure for a given application (El Madhoun et al., 2020) or whether a use case fulfills mandatory requirements and supportive characteristics for the appropriate use (Gallersdörfer & Matthes, 2020).

Specific to FedML, Bharti et al. (2022) proposed a decision model to capture the pathway of implementing a FedML ecosystem. Their artifact focus on technical implementation decisions but does not consider factors related to the business benefit and appropriate mapping of use case and technology for the technology selection. We aim to close this research gap by proposing a two-step artifact that aids in capturing the applicability and estimated technical complexity of their FedML project. Thereby, practitioners and decision-makers can assess the reasonable usage of FedML for a given use case and choose the appropriate technology.

3. Research Methodology

We observed in various focus group discussions that many FedML projects arise but fail to actualize due to an inappropriate fit between the use case and technology. Through a dedicated focus group discussion with four experts from two project teams, we intended to explore the problem space and understand the practitioners' perspectives. The focus group comprised a project lead, solution specialist, product manager, and ML engineer representing the target users. We recognized, that the novelty of FedML as well as the lack of best practices and guidelines pose the main challenges for practitioners. This makes it difficult to assess the appropriate use of FedML, estimate its implementation complexity, and thereby impede reasonable decisions about its adoption. Also, the scarcity of success stories of FedML usage makes organizations reluctant to invest.

Current literature does not offer guidelines on the appropriate use of FedML with regard to a given use case and its accompanying implementation complexity. Our study aims to close this research gap. We leveraged the Design Science Research (DSR) methodology as proposed by Peffers et al. (2007) to develop a tool that helps practitioners assess the reasonability and complexity of their planned FedML project. We chose the DSR approach since it provides a rigorous approach for producing and evaluating innovative, purposeful artifacts for a specific problem domain (Hevner et al., 2004). Table 1 provides an overview of our research approach and a short description of the conducted activities throughout the DSR cycle. As described in Section 4.2, we grounded our knowledge base in related work on AI adoption in organizations and technology selection tools, as well as on the influential factors of FedML projects (Müller et al., 2024).

4. Results

In the following, we will describe the results of our research. In accordance to the DSR methodology, we will start by outlining the identified *Objectives of a Solution*, followed by the *Design and Development* process and the *Demonstration and Evaluation*.

4.1. Objectives of a Solution

The objectives of a solution were identified through an initial focus group discussion and an expert interview study (Müller et al., 2024). Thereby, we aim to answer RQ1. We asked the four focus group participants as well as the 13 experts from the interview study what they require to assess the reasonability and complexity of a FedML project. Additionally, we discussed the design, content, and elements of a potential solution artifact.

Summarized, the participants require an artifact consisting of two components. The first component is an artifact that aids in determining the technology suitability, which includes assessing the benefits and disadvantages of FedML. This could help to determine whether FedML is a reasonable fit for the needs of the underlying use case and evaluates the added value of FedML. The second component

Table 1. Design Science Research steps according to Peffers et al. (2007).

Step	Short Description of Activities
(1) Problem Identification & Motivation	Identified the problem and motivation through focus group discussions. See the description above.
(2) Objectives of a Solution	Conducted focus group discussions and an interview study to derive requirements and determine relevant design principles. See Chapter 4.1 on Objectives of a Solution.
(3) Design & Development	Designed and developed the artifacts to provide decision-support and guidance for the initial complexity estimation for FedML projects. See Chapter 4.2 on Design and Development.
(4) Demonstration	Demonstrated the artifacts in group discussions and through usage of target users. See Chapter 4.3 on Demonstration and Evaluation.
(5) Evaluation	Evaluated the value, usability, efficiency, content, and structure of the artifacts. See Chapter 4.3 on Demonstration and Evaluation.
(6) Communication	Communication is being done through this paper.

deals with the expectable complexity of the FedML project due to the high variability of the components that have to be considered for the implementation. The more requirements are fulfilled and the less additional complexity is involved in the use case properties, the easier the implementation of the FedML project would be. In order to be able to quickly assess the suitability and complexity of the project implementation, the focus group suggests a checklist of requirements. The interviewees further requested a more sophisticated estimation of the expected complexity with the specific project circumstances. Especially with multiple implementation variations of FedML, the implementation efforts depend on the project's specifics. Therefore, an interactive artifact that captures the estimated complexity is required.

Consequently, we consider a two-step artifact consisting of a **Decision Tree** to assess the technology suitability for a use case and a subsequent **Survey** for the complexity estimation. The remainder of this paper will describe the requirements, design, development, and evaluation in general and separately for both artifacts.

Design Principles. We formulated a set of design principles to minimize errors, maintain high quality, and improve accountability. These principles apply to both artifacts and ensure consistency in the development process. Firstly and as highlighted by Hevner et al. (2004), the communication of DSR must address both, technology-oriented and management-oriented audiences. As a second design principle, the artifact needs to be broadly applicable in a variety of contexts while maintaining its relevance and validity regardless of the specific use case. Thirdly, we place significant emphasis on demonstrating the usefulness, quality, and effectiveness of the design artifact through rigorous evaluations (Hevner et al., 2004). This principle emphasizes the need for robust evaluation methods to objectively assess the effectiveness of the artifact in achieving its goals and to ensure that the artifact is credible and reliable. By complying with these design principles, we aim to maintain academic rigor, practical relevance, and broad applicability.

Specific Requirements to Decision Tree. The decision tree aims to critically examine the motivation to use FedML while highlighting the technology's potential advantages in order to challenge the technology decision. In addition to the design principles, the artifacts should provide examples of alternative technologies in case FedML is not the appropriate solution. Also, the different types of decisions should be color coded.

Specific Requirements to Survey. The survey should aid in identifying key aspects that need to be considered, particularly at the beginning of the technology selection process. The practitioners proposed that the questions are sorted from high-level to fine-grained questions and that the relevancy of the factors is taken into account. Also, the added complexity for each aspect should be indicated. Lastly, the suitability of the technology should also be underlined through color-codes. All requirements *R1-R7* are listed in Table 2.

4.2. Design and Development

To build our knowledge base, we identified and reviewed literature on the success factors of AI adoption in organizations (Alsheibani et al., 2019; Chatterjee et al., 2021; Dora et al., 2022; Hamm & Klesel, 2021; Kruse et al., 2019), technology selection tools (Friedrich

Table 2. Overview of identified requirements to the solution artifacts.

Artifact	ID	Short Description of Requirement
Both	R1	Understandable and usable by technology-oriented and management-oriented audiences
	R2	Generically applicable, not tailored to specific use cases
Decision Tree	R3	Provide examples of alternative technologies
	R4	Different types of decisions should be color-coded
Survey	R5	Sort questions from high-level to fine-grained
	R6	Impact on the project's complexity of each question/factor should be indicated
	R7	Complexity should be indicated through color-codes

et al., 2015; Maghazei et al., 2022; Shehabuddeen et al., 2006; Yap & Souder, 1993), and technology tools for other emerging technologies such as Blockchain (El Madhoun et al., 2020; Gallersdörfer & Matthes, 2020; Wüst & Gervais, 2018). The study is also based on an expert interview study on the influential factors of FedML projects (Müller et al., 2024). The study concluded that the most critical factors of FedML projects are technical considerations such as data quality and interoperability, as well as organizational and environmental aspects like organizational readiness, collaboration management, and legal regulations Müller et al., 2024. The findings from the interview study, focus group, and literature were incrementally combined to build the knowledge base. During development, we conducted regular focus groups with varying participants to iteratively assess the artifacts and implement feedback. The following aims to answer RQ2.

Decision Tree. The development of the decision tree followed the overall design principles and its above-mentioned specific requirements. The decision tree serves as a comprehensive tool for evaluating the strategic fit of FedML technology in the context of the specific use case and facilitates a quick assessment to identify cases where FedML may not be appropriate. Nevertheless, given the complexity of evaluating use cases, the artifact does not permit final decisions and requires further in-depth analysis. Therefore, the decision tree primarily provides recommendations, including situations where the underlying problem may not be solvable by FedML. The content of the decision tree is based on the value drivers and success factors of FedML which were identified through the literature review, focus group discussion, and interview study.

The structure of the decision tree is based on a binary approach, where each yes-no-question either leads to another question or to a recommendation. The decision tree ultimately contains eight questions and seven recommendations, including three cases where FedML seems not an appropriate solution, one case where FedML may not be necessary, one action for searching data collaborators, and three cases where FedML is considered likely to be appropriate. Decisions that advise against the usage of FedML are color-coded in red, the need for further actions is shown in blue, and decisions supporting FedML usage are indicated in green. Thereby, we fulfill *R4*. The decision tree also considers alternative technology options when FedML is not deemed appropriate. By presenting these alternative technology options, we intend to satisfy *R3*. The tree allows a subdivision by content, beginning with whether the use case is an ML task and whether the data must remain decentralized. From there, the necessity of the value drivers is queried. Starting with data privacy, followed by communication efficiency, and finally computational efficiency. An illustration of the decision tree is shown in Figure 2. The final decision tree can be accessed on the linked website[1].

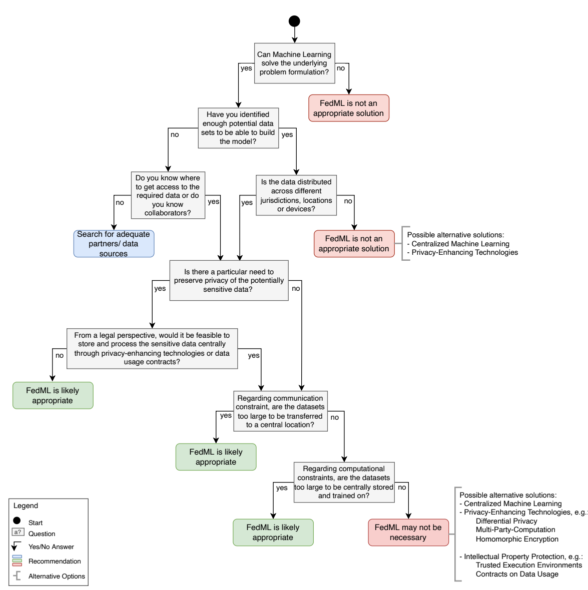

Figure 2. Design of the decision tree.

[1]Decision Tree and Survey: bit.ly/3EGcOFo

Survey. Based on the design principles and identified objectives, we developed a survey with questions on the influential factors of FedML projects. Each question in the survey is weighted according to its impact on the complexity of the implementation or its importance in the decision-making process. To meet *R7*, we color-coded the impact of answering each question. We selected Microsoft Excel[2] as the preferred tool for developing this interactive survey because of its ability to display information, accumulate factors, create interactive fields, conditional formatting, and its widespread use, which ensures good accessibility.

The survey and weighting of factors are based on the influential factors of FedML projects, which were identified in the literature review, focus group discussion, and interview study. The survey comprises 18 binary yes-no questions which are organized into four categories: ”Organizational Readiness”, ”Collaboration and Governance”, ”Data Considerations” and ”Technical Infrastructure”. Since the artifact only intends to highlight critical aspects and support the technology selection process, users are only expected to give answers to the best of their ability, without the need for clear and fully reasoned answers. The questions are intended to serve as a guideline and to point out critical aspects that should be examined thoroughly afterward in the decision process of using FedML.

Additionally, the survey includes a scoring system for supportive factors, decisive factors, and a complexity counter for negative answers. The relevancy and scoring of each factor are based on the interview study. By adding the complexity counter and weighing the factors, we aim to fulfill *R6*. We also structured the questions coming from high-level, broader questions at the beginning of the survey to more fine-grained, technical questions in the later sections. Therefore, we aim to realize *R5*.

Out of 18 questions, nine are decisive factors for FedML projects. Also, five questions increase the complexity counter by one, and four questions increase the counter by two. A color-coding scheme of green (supportive factor counter), yellow (complexity counter), and red (decisive factor counter) indicates the impact of each question. Hereby, *R7* is met. Upon answering a question, the respective counter is incremented automatically. At any time, the state of the counter is displayed to the user in the header of the survey. In addition, the maximum values are shown and provide a basis for classification.

A screenshot of the survey is shown in Figure 3. The entire and final survey can be accessed on the linked website[1]. It is important to note, that the final artifacts on the linked website include the suggestions and revisions described in the Chapter on *Demonstration and Evaluation*.

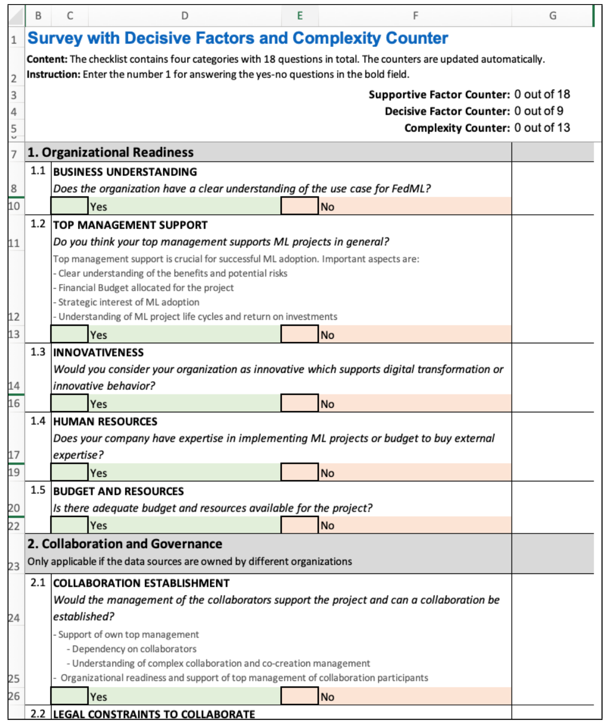

Figure 3. Design of the survey.

4.3. Demonstration and Evaluation

To demonstrate and evaluate the decision-support artifacts, we performed a demonstration with survey-based evaluations. The evaluation objective was to obtain quantitative and qualitative feedback from potential users and to assess whether key stakeholders perceive the solution artifacts as valuable. A total of 14 potential target users tested and evaluated the artifacts based on the perceived utility for their needs. Through an additional survey-based evaluation, we intended to measure how the solution artifacts represent a solution to the problem and also get an assessment of the perceived value (Prat et al., 2014). Via an additional blank field for open feedback, we also aimed to gather qualitative feedback besides quantitative feedback.

The evaluation criteria were selected in accordance with the requirements and objectives of the solution artifacts. To comply with the proven methods, we incorporated criteria specified by Prat et al. (2014). Consequently, we categorized a total of nine evaluation criteria regarding goal, content, and structure. The survey evaluates if the model is deemed valuable,

[2]https://www.microsoft.com/microsoft-365/excel

useful, and efficient. Moreover, it gathers estimations covering all important aspects, being detailed enough, comprehensible, and business-user friendly. Lastly, it is measured if it is well-structured and easy to use. The evaluation criteria were measured on a 5-point Likert scale.

Participant Selection. The participants for the demonstration and evaluation represent potential target users and include participants from the focus group, prior interviewees, and experts with no prior knowledge of the underlying research design and development. Thus, we can also assume a realistic and unbiased use of the solution artifacts. The evaluation group comprises a representative diverse range of backgrounds, needs, and views from eight organizations, including large enterprises, research centers, and start-ups. The full list of anonymized participants can be seen in Table 3. Overall, 14 participants from eight different companies used, tested and evaluated the artifacts.

Evaluation of Decision Tree. As displayed in Figure 4, all experts confirmed, that the decision tree is useful and delivers value. Therefore, we consider the goal of the decision tree to be reached. Apart from two neutral responses, the artifact is also perceived as efficient.

Most experts agree on the content of the decision tree. Four evaluators are missing some aspects and details regarding legal regulations and alternatives for privacy-enhancing technologies [EG-2, EG-1, EG-5, EG-14]. We included these aspects in our model and therefore assume that the final model now covers all important aspects. All except one expert [EG-5] perceived the artifact as comprehensible. The large majority of the experts strongly agree that the decision tree is business-user friendly. Lastly, the artifact is unanimously considered well-structured and easy to use.

The open feedback also expresses the strong approval from the experts. For example, one expert validated the content and style by stating:

> ”*I like the diagram. From the perspective of someone who builds FedML systems all the major questions asked are valid. And it's also delivered in a manner that makes it easily accessible to a non-technical audience.*” [EG-10]

Moreover, the purpose of the artifact was recognized and the decision tree was experienced as a ”*simple and effective tool to help managers make better-informed decisions*” [EG-3].

The open feedback also suggested improvements in wording, which we incorporated. Some experts

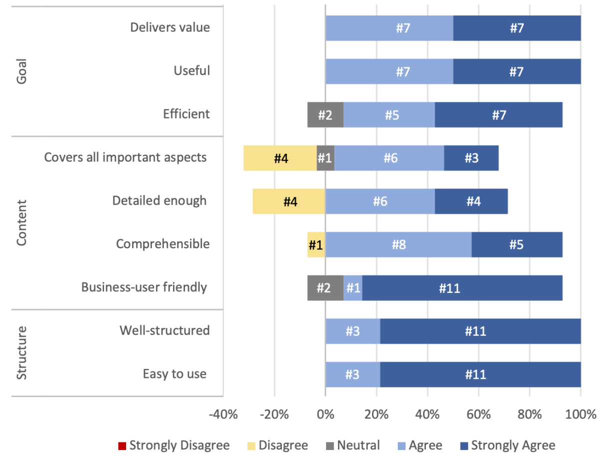

Figure 4. Evaluation results of the decision tree.

stated that it is difficult to answer the questions unambiguously. At this point, it is important to note, that the purpose of the artifact is a fast initial assessment of the technology's suitability. In case the technology seems to fit, it is still required to perform a subsequent thorough technology assessment.

The results show that the decision tree is understandable and usable by technology-oriented and management-oriented audiences. Also, since all experts from the broad evaluation group demographic and diverse backgrounds deem the model useful and efficient, we conclude that the decision tree is also generically applicable. Therefore, we consider requirements *R1* and *R2* for the decision tree as reached.

Evaluation of Survey. As displayed in Figure 5, the majority of experts validated that the survey is useful, efficient, and delivers value. One expert did not experience the survey as efficient and stated:

> ”*Most of these questions are much more complex than a simple yes/no. Nevertheless, the questionnaire gives orientation.*” [EG-5]

We want to highlight, that the survey only intends to provide orientation and guidance, not a thorough technical assessment. Therefore, we did not incorporate any changes since this is a conscious limitation of the artifact. Most experts confirm that the artifact contains all important aspects, is comprehensible, detailed enough, and business-user friendly.

Regarding completeness and level of detail, two experts [EG-1, EG-2] suggested visualizing the ”supportive” factors as well. One evaluator [EG-4] was missing legal constraints. Expert [EG-13] was missing a visual counter aggregating the positive results.

Table 3. Overview of experts for demonstration and evaluation.

ID	Position	Organization	Experience
EG-1	Product Manager	Large software enterprise 1	$\geq$ 10 years
EG-2	Architect	Large software enterprise 1	$\geq$ 5 years
EG-3	Applied Researcher	Industrial software enterprise	$\geq$ 2 years
EG-4	Development Expert	Large software enterprise 1	$\geq$ 19 years
EG-5	CEO and Founder	Startup for FedML 1	$\geq$ 3 years
EG-6	Applied Researcher	Research center for AI security	$\geq$ 1 year
EG-7	Senior Consultant and Project Lead	Large software enterprise 1	$\geq$ 6 year
EG-8	CEO and Founder	Startup for FedML 2	$\geq$ 5 years
EG-9	Researcher	Research center for software systems	$\geq$ 4 years
EG-10	Research Manager	Large software enterprise 2	$\geq$ 4 years
EG-11	Technology Consultant	Consultant company	$\geq$ 2 years
EG-12	Senior Researcher for Privacy	Large software enterprise 1	$\geq$ 7 years
EG-13	ML Engineer and Senior Data Scientist	Large software enterprise 1	$\geq$ 7 years
EG-14	Project Manager	Research center for AI security	$\geq$ 1 years

We included the suggestion of [EG-1, EG-2, EG-4, EG-13] to further improve its usefulness. Therefore, we consider the final artifact as complete.

Regarding comprehensibility, an expert disagreed since "*technical knowledge is required to answer questions from sections 3 and 4, so a combination of tech/business users is likely required to complete*" [EG-10]. We revisited the questions and concluded that these questions are critical to the completeness of the artifact and cannot be further abstracted due to the complex technical nature of the technology. Therefore, we kept sections 3 and 4. In general, the survey was also considered as well-structured and easy to use.

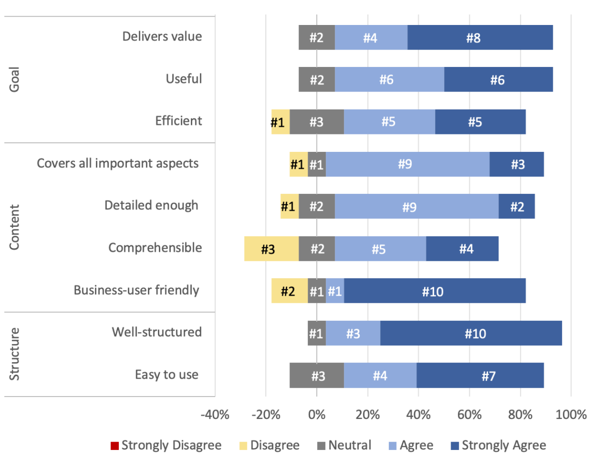

Figure 5. Evaluation results of the survey.

Overall, the structure of the artifact was perceived as "*very well designed*" [EG-11] and the open feedback again expresses strong approval from the experts. One expert additionally expressed that:

> "*This is also a very efficient tool for decision-makers to understand where the potential problems may arise. If anything, the descriptions could be a bit longer but in general, I think it achieves its purpose.*" [EG-3]

Similar to the feedback on the decision tree, we received suggestions to improve the wording of some questions [EG-10, EG-1, EG-12]. We included the feedback by eliminating double negatives and using less technological vocabulary.

Overall, the evaluation results suggest that technology-oriented as well as management-oriented audiences understand the artifact and deem the survey useful. Additionally, the results show that the survey delivers value to all experts from diverse backgrounds and with diverse requirements. Therefore, we consider requirements *R1* and *R2* as fulfilled for the survey. Through this, we also achieved to meet all pre-defined requirements for the artifacts.

Through the feedback we gathered in the demonstration and evaluation, we were able to further improve the artifacts to the requirements and needs of the target users and thereby strengthen our contribution.

5. Discussion

We recognized that practitioners lack guidance in assessing the suitability and complexity of implementing FedML in their use case. The problem was identified and motivated through a focus group discussion with four potential users from two project teams. In addition, the relevancy was further confirmed by an interview study and in the evaluation. We developed the decision-support tool according to pre-defined design principles and solution requirements from practitioners. Moreover, we refined the artifacts through feedback from the evaluators. As outlined in Chapter 4.2 and Chapter 4.3, the final decision-support tool meets all requirements for the solution artifacts.

Contributions. Current literature mostly takes a technical perspective on the design of FedML systems and lacks a business perspective on technology adoption. With our study, we extend current information systems literature on FedML by a methodology for investigating the reasonable adoption of FedML. We propose an artifact design to provide guidance for practitioners and decision-makers in the technology selection. This helps practitioners understand the implications of adopting FedML and hereby, contribute to a well-grounded basis for decision-making. Our artifacts provide guidance for technology-oriented users as well as users with a managerial perspective. By highlighting potential critical factors through our decision-support tool, we facilitate the development of FedML projects and thereby support the broad practical adoption of FedML.

Limitations. As with any study, our research has limitations. DSR involves iterative design and development cycles which may be affected by the subjectivity of the researchers. To address this risk, two researchers participated in the focus groups and interviews to ensure observer triangulation. We established design principles for the development process and all results were iteratively double-checked by an additional researcher. Moreover, it must be noted that the artifacts only intend to provide orientation and guidance. The yes-no questions are not meant as the sole basis for decision-making, nor do they expect a clear response from the practitioners. Given the limited number of experts, this research is exploratory and requires further confirmation with practical validation. Since the models could not be directly tested in case studies, we convened evaluation participants without prior knowledge of our research. Thus, these experts represent potential target users with a realistic usage environment. However, due to the limited number of experts, we encourage practitioners to test our artifacts in case studies to further develop results. Due to the novelty of the technology, we were only able to survey early adopters for our knowledge base. Therefore, the results could change as the technology is adopted more broadly. Other factors could emerge and some could be mitigated by the emergence of best practices or a change in business understanding towards FedML.

6. Conclusion

In this study, we introduced a two-step decision-support tool to provide practitioners with an evidence-based support tool to aid well-grounded technology selection decisions for FedML projects. The artifacts complement current research on FedML systems and the corpus of information systems literature on technology selection. Through this, we contribute to business model innovation by facilitating the successful adoption of FedML, and consequently the creation of substantive value through collaborative AI projects. The evaluation results show that potential users deem the tool to be useful, deliver value, and be efficient. Furthermore, the evaluation group validated the content and structure of the artifacts. We showed that the tool can help the target users in a well-grounded assessment of the suitability and expected complexity of the FedML usage in their use case. We encourage practitioners to test our artifacts in case studies to further develop the artifacts, use our insights as a basis for future research, and adapt the tool with potentially emerging best practices for the adoption of FedML in organizations.

7. Acknowledgements

The authors would like to thank SAP SE for supporting this work.

References

Alsheibani, S., Cheung, Y., & Messom, C. (2018). Artificial intelligence adoption: AI-readiness at firm-level. *Pacific-Asia Conference on Information Systems*, *37*.

Alsheibani, S., Cheung, Y., & Messom, C. (2019). Factors inhibiting the adoption of artificial intelligence at organizational-level: A preliminary investigation. *AMCIS 2019 Proceedings. 2*.

Bauer, M., van Dinther, C., & Kiefer, D. (2020). Machine learning in SME: An empirical study on enablers and success factors. *AMCIS 2020 Proceedings 3*.

Bharti, S., McGibney, A., & O'gorman, T. (2022). Design considerations and guidelines for implementing federated learning in smart manufacturing applications. *IIC Journal of Innovation*, *19*, 17–35.

Chatterjee, S., Rana, N., Dwivedi, Y., & Baabdullah, A. (2021). Understanding AI adoption in manufacturing and production firms using an integrated TAM-TOE model. *Technological Forecasting and Social Change*, *170*(5), 34.

Dora, M., Kumar, A., Mangla, S. K., Pant, A., & Kamal, M. M. (2022). Critical success factors influencing artificial intelligence adoption in food supply chains. *International Journal of Production Research*, *60*(14), 4621–4640.

El Madhoun, N., Hatin, J., & Bertin, E. (2020). A decision tree for building IT applications: What to choose: Blockchain or classical systems? *Annals of Telecommunications*, *76*.

Friedrich, T., Overhage, S., Schlauderer, S., & Eggs, H. (2015). Selecting technologies for social commerce: Towards a systematic method. *ECIS 2015*.

Gallersdörfer, U., & Matthes, F. (2020). Towards valid use cases: Requirements and supporting characteristics of proper blockchain applications. *2020 Seventh International Conference on Software Defined Systems (SDS)*, 202–207.

Giray, G. (2021). A software engineering perspective on engineering machine learning systems: State of the art and challenges. *Journal of Systems and Software*, *180*, 111031.

Hamm, P., & Klesel, M. (2021). Success factors for the adoption of artificial intelligence in organizations: A literature review. *Proceedings of the 27th Americas Conference on Information Systems*, *1*.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.

Hofmann, P., Jöhnk, J., Protschky, D., & Urbach, N. (2020). Developing purposeful AI use cases - A structured method and its application in project management. *Entwicklungen, Chancen und Herausforderungen der Digitalisierung: Proceedings der 15. Internationalen Tagung Wirtschaftsinformatik*, 33–49.

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes — An interview study of organizational AI readiness factors. *Business & Information Systems Engineering*, *63*, 1–16.

Kruse, L., Wunderlich, N., & Beck, R. (2019). Artificial intelligence for the financial services industry: What challenges organizations to succeed. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 6408–6417.

Lo, S. K., Lu, Q., Wang, C., Paik, H.-Y., & Zhu, L. (2021). A systematic literature review on federated machine learning: From a software engineering perspective. *ACM Comput. Surv.*, *54*(5).

Maghazei, O., Lewis, M. A., & Netland, T. H. (2022). Emerging technologies and the use case: A multi-year study of drone adoption. *Journal of Operations Management*, *68*(6-7), 560–591.

McMahan, H. B., Moore, E., Ramage, D., & y Arcas, B. A. (2016). Federated learning of deep networks using model averaging. *ArXiv*.

Müller, T., Zahn, M., & Matthes, F. (2023). Unlocking the potential of collaborative AI - On the socio-technical challenges of federated machine learning. *ECIS 2023*.

Müller, T., Zahn, M., & Matthes, F. (2024). Revealing the impacting factors for the adoption of federated machine learning in organizations. *HICSS 2024*.

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*, 45–77.

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). Artifact evaluation in information systems design-science research – A holistic view. *PACIS 2014*.

Schomakers, E.-M., Lidynia, C., & Ziefle, M. (2020). All of me? Users' preferences for privacy-preserving data markets and the importance of anonymity. *Electronic Markets*, *30*.

Shehabuddeen, N., Probert, D., & Phaal, R. (2006). From theory to practice: Challenges in operationalising a technology selection framework. *Technovation*, *26*(3), 324–335.

Shen, Y.-C., Chang, S.-H., Lin, G. T., & Yu, H.-C. (2010). A hybrid selection model for emerging technology. *Technological Forecasting and Social Change*, *77*(1), 151–166.

Vorraber, W., Neubacher, D., Mösl, B., Brugger, J., Stadlmeier, S., & Voessner, S. (2019). UCTM - An ambidextrous service innovation framework - A bottom-up approach to combine human- and technology-centered service design. *Systems*, *7*, 23.

World Economic Forum. (2020). Share to gain: Unlocking data value in manufacturing. https://www3.weforum.org/docs/WEF_Share_to_Gain_Report.pdf

Wüst, K., & Gervais, A. (2018). Do you need a blockchain? *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, 45–54.

Yap, C. M., & Souder, W. E. (1993). A filter system for technology evaluation and selection. *Technovation*, *13*(7), 449–469.