Selecting Adequate Machine Learning Methods for Human-Computer Interaction Data Sets: Guidelines and a Conceptual Structure

Anna C. Weigand[1,2,*]

[1]*University of Seville, Spain*

[2]*University of Applied Sciences Emden/Leer, Germany*

Abstract

Data sets in human-computer interaction (HCI) research vary widely in their characteristics. Due to the methods of data collection, such as usability testing and interviews, HCI data sets are usually smaller than big data sets, such as those of social media and image processing. However, HCI data sets can also be complex and multi-dimensional, which makes pattern recognition difficult. Therefore, we investigate the impact factors and parameters to consider when applying machine learning (ML) methods to HCI data sets. Through iterative structured and unstructured expert interviews and machine learning experiments, we identify impact factors and parameters that lead to a conceptual structure. Overall, our aim is to deduce and provide decision-making support for selecting an adequate ML method for HCI data sets.

1. Motivation

In the research field of human-computer interaction (HCI) [1], a wide variety of HCI data sets is available. Methods of data collection include usability tests, interviews, focus groups, and surveys [2]. Data can also be extracted from large-scale user studies [3, 4], eye-tracking experiments [5], and web tracking [6].

Although all of these data sets are HCI data sets, they have varying characteristics [3, 5, 7, 8]. The most apparent one is data set volume, which tends to be small in HCI research (in contrast to big data sets from, *e.g.,* social media) and typically includes only about 200 data points [9, 10]. In addition, data are collected in a variety of contexts with varying goals or purposes [11]. Data are also generated by different users, which can lead to biases [12] and variances [8], *e.g.,* due to imbalanced data sets. Furthermore, since HCI data is primarily collected from and about users, it must be handled with care, both in terms of management and outcomes. Nevertheless, HCI data sets seem to be influenced by similar impact factors, such as the *individuals* who are the basis for data gathering (*e.g.,* in terms of participation in interviews or usability tests) or *regulations* concerning data privacy. Some of these factors arise from certain standards, such

as those provided by the human-centered design (HCD) process model as per ISO standard 9241-220 [13].

As HCI data are not necessarily big data but can be complex and multi-dimensional regardless of size, patterns may not be detected by standard data analysis methods, as illustrated by the Kano categorization of about 400 participants' survey answers [3]. Therefore, machine learning (ML) can help researchers extract further insights from highly qualitative HCI data with less manual effort and perhaps better results.

The role of ML in HCI research has already been discussed in the literature. However, a survey showed that the application of ML to HCI data sets is challenging and that appropriate process models are needed to proceed adequately [14]. Another study summarized some pitfalls to avoid when using ML for HCI research data, such as "*classification accuracy is not hypothesis testing*" and "*causality versus correlation*" [15]. Furthermore, according to the authors, referencing classifiers' baseline performance ensures the transparency of the results.

However, to the best of our knowledge, applying ML to HCI data sets has not yet been systematically investigated. This raises the fundamental question: *How can one select an adequate approach and ML methods for HCI data sets?* To answer this research question, we build on our research approach described below.

2. Research Approach

As a basis for our research, we use the design science research methodology (DSRM) [16]. Figure 1 shows the particular steps of the DSRM as well as the related goals to be achieved.

First, we *identify the problem* through a literature review [10]. We then set the *objectives* for solving the problem through our study concept. To *design and develop* a suitable solution to our problem, we conduct structured and unstructured interviews with HCI and ML experts to identify relevant aspects (i.e., those that affect the selection of ML methods for HCI data sets). In addition, we expect to gain a holistic view of these aspects through ML experiments on HCI datasets. To *demonstrate* the solution to our problem, we aim to develop a conceptual structure (CS) based on our insights from the interviews and ML experiments that depicts various impact
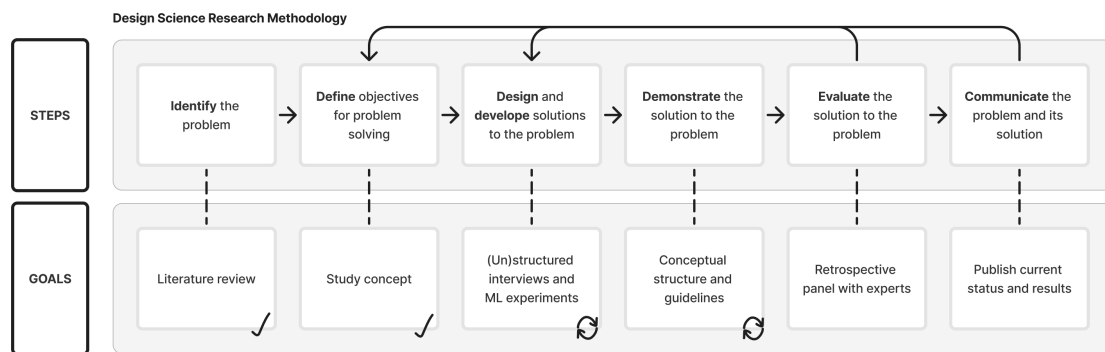

Design Science Research Methodology

| STEPS | **Identify** the problem | **Define** objectives for problem solving | **Design** and **develope** solutions to the problem | **Demonstrate** the solution to the problem | **Evaluate** the solution to the problem | **Communicate** the problem and its solution |

| GOALS | Literature review ✓ | Study concept ✓ | (Un)structured interviews and ML experiments ↻ | Conceptual structure and guidelines ↻ | Retrospective panel with experts | Publish current status and results |

Figure 1: The design science research methodology [16] applied to this work. The check marks show which goals have been achieved so far, and the iteration icon indicates which ones are still in development.

factors and their parameters to consider when selecting an ML method for HCI data sets. The CS is meant to represent the activities of all parties involved, show the various interdependencies between all of the impact factors (similar to a metamodel or conceptual model [17, 18]), and facilitate decisions for specific ML methods. Furthermore, we intend to provide guidelines for HCI analysis with ML and the selection of ML methods for HCI data sets. Subsequently, we conduct a retrospective expert panel to *evaluate* our results. Regular publication of our current status and results guarantees *communication*.

In addition, we take advantage of an iterative and agile development process, as suggested by the well-established Agile Manifesto, to create the CS [19, 20]. This is an appropriate and adjustable approach that allows us to react quickly to new developments, which is critical because the ML field has been growing rapidly. For our studies, we focus on the exchanges with experts and their experiences in the fields of ML and HCI, as well as on our project experiences. We aim to invent a realistic and applicable CS that is continually updated according to new findings (see Figure 2).
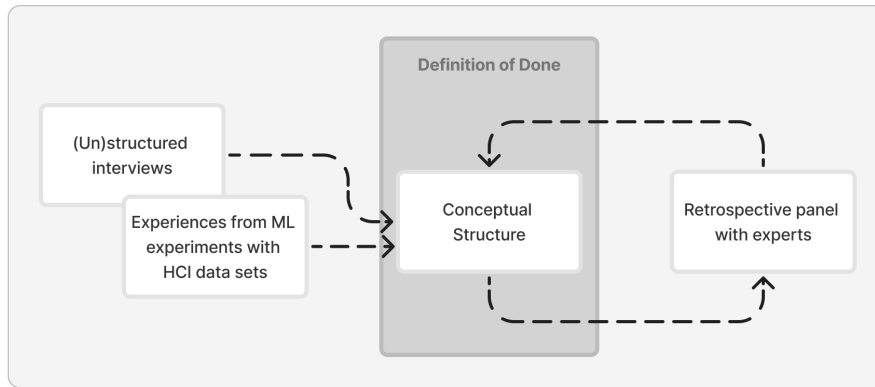

Figure 2: Our iterative approach for developing a valid conceptual structure.

The starting point is our first draft of the CS (see Figure 3), which is iteratively improved by structured expert interviews. We also plan to conduct ML experiments with HCI data sets to enrich our CS with further practical knowledge.

Specifically, we apply ML methods to HCI data sets of different use cases (*e.g.,* health-related data sets collected with an online experiment [4]) to deduce new findings for our CS. Furthermore, as established in the agile software development approach of the Scrum Guide [21], we plan a *retrospective* expert panel with stakeholders from industry and science to discuss the current CS, identify further impact factors and parameters, and gather additional feedback. To define the end of the iterations for the CS, we use the agile practice *definition of done*.

3. Preliminary Results

In the first step, we conducted ML experiments with prediction methods [8, 22] and clustering methods [23] on health-related data sets (around 300 data points) collected with an online experiment to gain experience with the process.
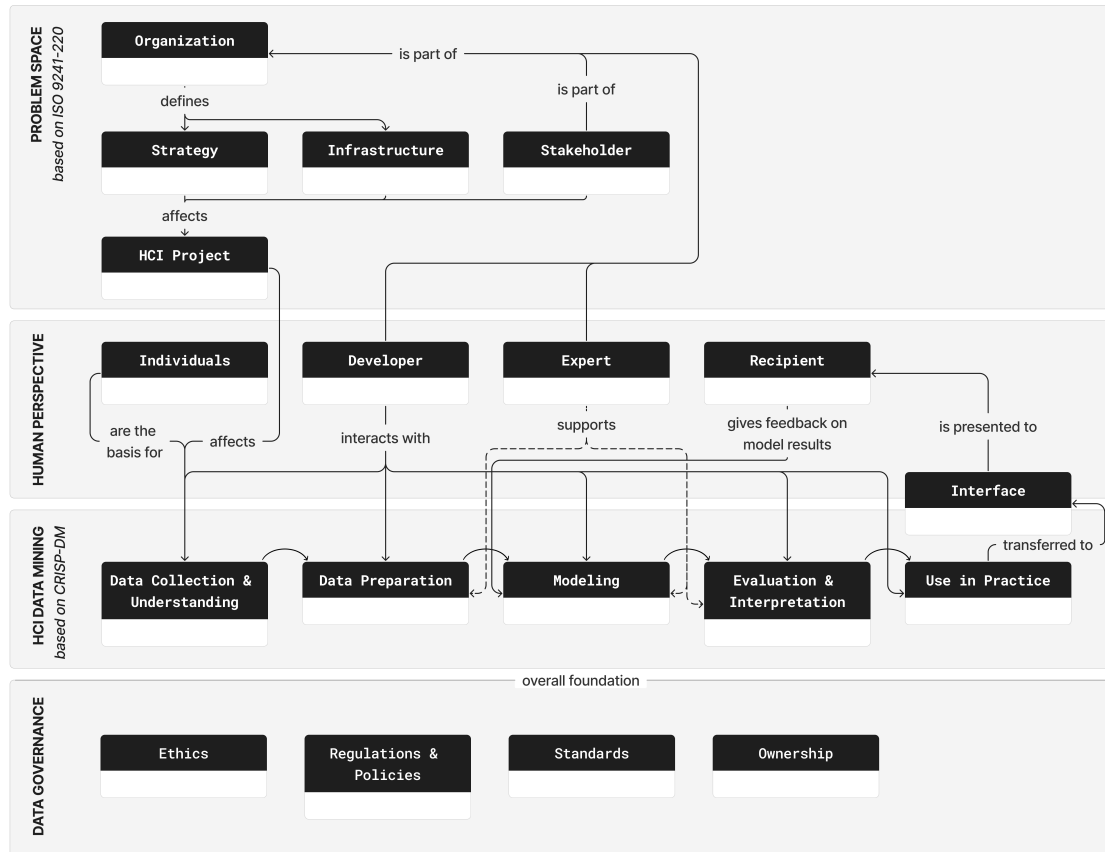
Figure 3: First draft and simplified version of our conceptual structure, which is organized into four horizontal dimensions and includes impact factors with their related parameters (parameters are not shown in this simplified version).

We have also done a systematic literature review [23] to define the context of small HCI data sets [10] that are analyzed with ML techniques in the HCI community. Since the data sets of the remaining articles were often not sufficiently described in this context, we did a first rough analysis of *n* = 29 articles, which showed us that mainly data sets with less than 200 data points were used as input for ML.

Our first draft of a CS (see Figure 3) consolidates the findings from our previous research, three informal, unstructured interviews with industry experts, the author's opinion, and further feedback of five reviewers. Based on these insights, four main topics arose. Therefore, the CS is divided into the following four dimensions: *problem space*, *human perspective*, *HCI data mining*, and *data governance*.

The first dimension, *problem space*, is based on the ISO standard 9241-220 [13]. This standard already provides an overview of the *problem space*, as it describes an overall concept for implementing an HCD approach in organizations. The four categories (strategy, organizational infrastructure, project, and operation) of the HCD process model [13] were integrated into this first dimension of the CS, along with the relevant describing parameters. The interviews showed

us that two additional describing parameters are relevant for the impact factor *HCI project*: *objectives* and *project stage*. Therefore, we reduced the complexity of the CS in this dimension by adding the new parameter *project stage* to the impact factor *HCI project* and eliminating an extra impact factor called *operation*.

The second dimension, *human perspective*, consists of the impact factors *individuals*, *developer*, *expert* [24], and the *recipient* of the results. All of these parties influence the data mining process in their own way, taking the HCI project to the next level. For example, *individuals* are the basis for HCI data collection because their interaction with a system or its evaluation generates the HCI data.

The third dimension, *HCI data mining*, is based on the widely-used cross-industry standard process for data mining (CRISP-DM) [25]. Adapted variants of this standard for specific use cases also exist (*e.g.,* for manufacturing small and medium-sized enterprises [26]). In the context of this CS, however, the original CRISP-DM is used because we focus on HCI data sets rather than a specific field of application. To the best of our knowledge, there is not yet a specific standard process for this context. Therefore, we extract the following dimensions from CRISP-DM: data understanding, data preparation, modeling, evaluation, and deployment. In our CS, we modified the description for better understanding in the HCI context: *data collection & understanding*, *data preparation*, *modeling*, *evaluation & interpretation*, and *use in practice*. However, the first step of CRISP-DM is business understanding, which is not named separately in our CS but is merged with the impact factor *HCI project*. Between the impact factor *use in practice* and the *recipient*, we also added an *interface*, which allows the user to use the ML model or the results. This can be in the form of an application, but it can also be a presentation or something similar.

The fourth dimension, *data governance*, results from our unstructured interviews. The interview participants provided additional impact factors and parameters to consider for ML and AI applications. We decided to group them under the *data governance* dimension in our CS. Decision rights and responsibilities, data policies, and standards and compliance aspects all fall under the category of *data governance* [27]. For now, we also include ethical impact factors in the dimension *data governance* [27, 28]. However, according to most recent research, an ethics strategy should be defined first [29]. In our future work, this topic will be considered in more detail.

4. Research Plan and Next Steps

Based on our preliminary findings, we intend to further evaluate and verify our CS draft. We first plan to conduct additional structured interviews with HCI and ML experts to enhance the insights from previous interviews and acquire information that is still missing. We will then adapt the CS according to the new findings from these interviews (*e.g.,* by adding new impact factors or parameters or rearranging the existing structure). This will clarify the process and impact factors regarding ML application in the context of HCI research. In addition, during a *retrospective* panel with experts, we will evaluate our CS and deduce guidelines for ML usage on HCI data sets to validate our approach.

To execute this plan, we first need to prepare interview guidelines for our structured interviews with experts.

Furthermore, we will investigate the results of applying various ML algorithms to a variety of HCI data sets. The experiences gained from the experiments will both enrich our CS and reveal which attributes, methods, and approaches we need to consider to avoid bias in our results.

We also prepare relevant HCI data sets for our planned experiments as well as a *definition of done* for our CS.

Acknowledgment

References

[1] International Organization for Standardization, Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems, 2019.

[2] J. Lazar, J. H. Feng, H. Hochheiser, Research Methods in Human-Computer Interaction, Elsevier Inc., Cambridge, United States, 2017. doi:10.1016/b978-0-444-70536-5.50047-6.

[3] J. Deutschländer, A. C. Weigand, A. M. Klein, D. Winter, M. Rauschenberger, There are no major age effects for UX aspects of voice user interfaces using the Kano categorization, in: Proceedings of the 19th International Conference on Web Information Systems and Technologies, SCITEPRESS - Science and Technology Publications, 2023, pp. 330–339. doi:10.5220/0012187600003584.

[4] M. Rauschenberger, R. Baeza-Yates, L. Rello, A Universal Screening Tool for Dyslexia by a Web-Game and Machine Learning, Frontiers in Computer Science 3 (2022) 111. URL: https://www.frontiersin.org/article/10.3389/fcomp.2021.628634. doi:10.3389/fcomp.2021.628634.

[5] S. J. Garbin, O. Komogortsev, R. Cavin, G. Hughes, Y. Shen, I. Schuetz, S. S. Talathi, Dataset for eye tracking on a virtual reality platform, in: A. Bulling, A. Huckauf, E. Jain, R. Radach, D. Weiskopf (Eds.), ACM Symposium on Eye Tracking Research and Applications, ACM, New York, NY, USA, 2020, pp. 1–10. doi:10.1145/3379155.3391317.

[6] S. Dambra, I. Sanchez-Rola, L. Bilge, D. Balzarotti, When sally met trackers: Web tracking from the users' perspective, in: 31st USENIX Security Symposium (USENIX Security 22), USENIX Association, Boston, MA, 2022, pp. 2189–2206. URL: https://www.usenix.org/conference/usenixsecurity22/presentation/dambra.

[7] O. R. Ogunseiju, N. Gonsalves, A. A. Akanmu, D. Bairaktarova, D. A. Bowman, F. Jazizadeh, Mixed reality environment for learning sensing technology applications in construction: A usability study, Advanced Engineering Informatics 53 (2022) 101637. doi:10.1016/j.aei.2022.101637.

[8] M. Rauschenberger, R. Baeza-Yates, How to handle health-related small imbalanced data in machine learning?, i-com 19 (2020) 215–226. doi:`10.1515/icom-2020-0018`.

[9] K. Caine, Local standards for sample size at chi, in: J. Kaye, A. Druin, C. Lampe, D. Morris, J. P. Hourcade (Eds.), Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, 2016, pp. 981–992. doi:`10.1145/2858036.2858498`.

[10] A. C. Weigand, M. Rauschenberger, Exploring the definition of small data collected with HCI methods and used for ML, Mensch und Computer 2023 - Workshopband, 2023. doi:`10.18420/muc2023-mci-ws16-399`.

[11] R. Baeza-Yates, Big, small or right data: Which is the proper focus?, 2018. URL: https://www.kdnuggets.com/2018/10/big-small-right-data.html.

[12] A. S. Rohani, R. Baeza-Yates, Measuring bias, in: 2023 IEEE International Conference on Big Data (BigData), IEEE, 2023, pp. 1289–1298. doi:`10.1109/BigData59044.2023.10386679`.

[13] DIN Deutsches Institut für Normung, Ergonomie der Mensch-System-Interaktion – Teil 220: Prozesse zur Ermöglichung, Durchführung und Bewertung menschzentrierter Gestaltung für interaktive Systeme in Hersteller- und Betreiberorganisationen (iso 9241-220:2019); Deutsche Fassung EN ISO 9241-220:2019, Juli 2020.

[14] V. S. Moustakis, J. Herrmann, Where do machine learning and human-computer interaction meet?, Applied Artificial Intelligence 11 (1997) 595–609. doi:`10.1080/088395197117948`.

[15] V. Kostakos, M. Musolesi, Avoiding pitfalls when using machine learning in HCI studies, Interactions 24 (2017) 34–37. doi:`10.1145/3085556`.

[16] K. Peffers, T. Tuunanen, M. A. Rothenberger, S. Chatterjee, A design science research methodology for information systems research, Journal of Management Information Systems 24 (2007) 45–77. doi:`10.2753/MIS0742-1222240302`.

[17] M. J. Escalona, N. Koch, Metamodeling the requirements of web systems, in: J. Filipe, J. Cordeiro, V. Pedrosa (Eds.), Web Information Systems and Technologies, Springer eBook Collection Computer Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 267–280. doi:`10.1007/978-3-540-74063-6{\textunderscore}21`.

[18] E.-M. Schön, M. Neumann, C. Hofmann-Stölting, R. Baeza-Yates, M. Rauschenberger, How are AI assistants changing higher education?, Frontiers in Computer Science 5 (2023). doi:`10.3389/fcomp.2023.1208550`.

[19] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. Martin, S. Mellor, K. Schwaber, J. Sutherland, D. Thomas, Manifesto for agile software development, 2001. URL: https://agilemanifesto.org/.

[20] E.-M. Schön, J. Thomaschewski, M. J. Escalona, Agile requirements engineering: A systematic literature review, Computer Standards & Interfaces 49 (2017) 79–91. doi:`10.1016/j.csi.2016.08.011`.

[21] K. Schwaber, J. Sutherland, The Scrum Guide, 2020. URL: https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-US.pdf.

[22] M. Rauschenberger, R. Baeza-Yates, Recommendations to handle health-related small imbalanced data in machine learning, Mensch und Computer 2020 - Workshopband, 2020.

doi:`10.18420/muc2020-ws111-333`.

[23] A. C. Weigand, D. Lange, M. Rauschenberger, How can small data sets be clustered?, Mensch und Computer 2021 - Workshopband, 2021. doi:`10.18420/muc2021-mci-ws02-284`.

[24] X. Wu, L. Xiao, Y. Sun, J. Zhang, T. Ma, L. He, A survey of human-in-the-loop for machine learning, Future Generation Computer Systems 135 (2022) 364–381. doi:`10.1016/j.future.2022.05.014`.

[25] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, CRISP-DM 1.0: Step-by-step data mining guide, 2000. URL: https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf.

[26] S. Rösl, T. Auer, C. Schieder, Addressing the data challenge in manufacturing SMEs: A comparative study of data analytics applications with a simplified reference model, in: M. Elstermann, A. Dittmar, M. Lederer (Eds.), Subject-Oriented Business Process Management. Models for Designing Digital Transformations, volume 1867 of *Communications in Computer and Information Science*, Springer Nature Switzerland, Cham, 2023, pp. 121–130. doi:`10.1007/978-3-031-40213-5_9`.

[27] R. Abraham, J. Schneider, J. vom Brocke, Data governance: A conceptual framework, structured review, and research agenda, International Journal of Information Management 49 (2019) 424–438. doi:`10.1016/j.ijinfomgt.2019.07.008`.

[28] M. Janssen, P. Brous, E. Estevez, L. S. Barbosa, T. Janowski, Data governance: Organizing data for trustworthy artificial intelligence, Government Information Quarterly 37 (2020) 1–8. doi:`10.1016/j.giq.2020.101493`.

[29] R. Baeza-Yates, U. M. Fayyad, Responsible AI: An urgent mandate, IEEE Intelligent Systems 39 (2024) 12–17. doi:`10.1109/MIS.2023.3343488`.