



HAL
open science

Processing graphs as an illustration of how engineering students build a machine learning model

Katharina Bata, Angela Schmitz, Andreas Eichler

► **To cite this version:**

Katharina Bata, Angela Schmitz, Andreas Eichler. Processing graphs as an illustration of how engineering students build a machine learning model. Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13), Alfréd Rényi Institute of Mathematics; Eötvös Loránd University of Budapest, Jul 2023, Budapest, Hungary. hal-04410716

HAL Id: hal-04410716

<https://hal.science/hal-04410716>

Submitted on 22 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Processing graphs as an illustration of how engineering students build a machine learning model

Katharina Bata¹, Angela Schmitz¹ and Andreas Eichler²

¹TH Köln – University of Applied Sciences, Germany; katharina.bata@th-koeln.de;

²Universität Kassel, Germany

Data science and machine learning methods find increasing application in different fields of engineering. Model building is one important learning objective in the context of machine learning. In this paper, we analyse the structure of engineering students' first model building processes and identify five characteristic structures. In this context, a method for illustrating and analysing a model building process is introduced, the "processing graph". The processing graph and the characteristic structures provide the opportunity to describe the structure of individual model building processes and to connect these with content-related aspects in future research projects.

Keywords: Data Science, Machine Learning, Model Building, Processing Graph.

Introduction

As a result of the growing popularity and relevance of data science methods, especially machine learning (ML) in industry, research, and private life, education in this fields is gaining importance (Engel, 2017; Touretzki et al., 2019). However, teaching and learning ML is still an underexplored topic, and there are many unanswered questions regarding learning objectives, teaching methods, and learning processes (Steinbach et al., 2020). Two commonly accepted learning objectives, however, are model building and validation (Lavesson, 2010; Touretzki et al., 2019; Steinbach et al., 2020). Model building refers to applying a ML method to a data set to systematically explore, represent and use the information presented in the data. Model validation refers to checking the quality and applicability of a model.

The project upon which this paper is based follows the method of design research to combine research and development of teaching-learning-material for the introduction of ML methods for engineering students (Bata et al., 2022). In the context of this project, videos of students working through the developed material in a design experiment (Gravemeijer & Cobb, 2006) were recorded. This paper shows the results of a qualitative analysis of those parts of the videos, where the students work on a model building task. As a part of the analysis, the students' approaches are visualized with a so-called "processing graph", a method that we developed and which will be introduced in the methods section.

Theoretical background and research question

Teaching and learning of machine learning

Taking the number of published papers and instructional units for ML as a benchmark, the interest in teaching ML from early school years onwards has increased sharply in the last years (Marques et al., 2020; Bilstrup et al. 2022). The concretization of curricula for different target groups arises (Touretzki et al., 2019), and the amount of published best-practice concepts over the last years is as diverse as it

is versatile (Fiebrink, 2019; Huppenkothen & Eadie, 2020; Kinnaird, 2020). One of the most fundamental differences in the chosen teaching approaches is how the methodological foundations, such as mathematics or programming, are handled. Opinions differ on whether teachers should teach and use the methods as a black-box, or uncover the statistical and mathematical foundations of the methods (Biehler & Schulte, 2018, Bilstrup et al. 2022).

In the context of model building and model validation the statistical tests and performance measures seem to be hurdles for learners (Lavesson, 2010), which leads to approaches how to exclude these topics at all (Fiebrink, 2019). It is also unclear to what extent the use of a black-box or a white-box approach has an influence on the learners' processes, for example, when they build their own models (Heuer et al., 2010). To approach the model building processes, in previous research-steps we developed a category system (Bata et al., 2021). We used the categories to structure the processes according to the decisions made by the students and distinguished between justified and less justified decisions. This initial analysis of the processes was supplemented by a content analysis of the justifications which the students used for their decisions (Bata et al., 2023).

Research question

Building on the results of Bata et al. (2021, 2023), this paper aims to address the research question:

“What characteristic structures can be found in the students' model building processes?”

As a characteristic structure we define a phase or property occurring in the students' approaches, which can be found in more than one of the observed groups. We use the notion “structure” because the analysis rather covers the process than the content of the students' approaches. For example, it is considered whether a justification is correct in content or not, but not which content was used.

Method

The observed task

The analysis of the students' model building processes was conducted with a task, in which students work with a data set from a survey of the quality of steel, consisting of two metric features (number of cracks / width of the biggest crack) and a label (good quality / bad quality). The analyzed video material originates from design experiments of a design research project with twelve groups (Bata et al., 2022), where two to three students work on the task together in an online conference tool on a split screen with a Jupyter Notebook. The Jupyter Notebook contains the data set and some prepared cells with the basic framework for model building with k-Nearest Neighbor (kNN). There are also cells that can be used to display a selection of performance measures, and a graphical display of the model once with the training data and once with the test data. The students are asked to build a model on the dataset, and then to classify a new dataset using their model.

From analyzing with the category system to the processing graph

In Bata et al. (2021) it is described how the students' model building processes can be analyzed and structured with the category system shown in Table 1. In the first column there are the two main categories *activity* and *topic* (note: the third main category, which occurred in the publication in 2021, is not relevant to the current research question), in the second column there are all subcategories

(note: some subcategory names have changed slightly since their first publication in 2021). The third column is for additional codes, which do not fit into the logic of the subcategories but which are, as we will see later, important for results concerning the research question:

Table 1: Category system

| Main category | Subcategory | Additional codes |
|---------------|----------------------------------------|--------------------------------------------------------------------------|
| C1: Activity | C1.1: Decision, without content reason | C1*: Talking about systematic parameter optimization (e.g. using a loop) |
| | C1.2: Decision, targeted trying | |
| C2: Topic | C2.1: Testsize (t) | C2.4: Model |
| | C2.2: Number of neighbors (k) | C2.5: Others |
| | C2.3: Metric (p) | |

When analyzing the videos with the category system, between 15 and 40 coded passages were generated per video. In order to make the results of the coding and thus the students' approaches visible at a glance, we developed the "processing graph" as a visualization method. In the following, we show how such a processing graph represents the codes we applied within the model building process of group 6 (see Figure 1).

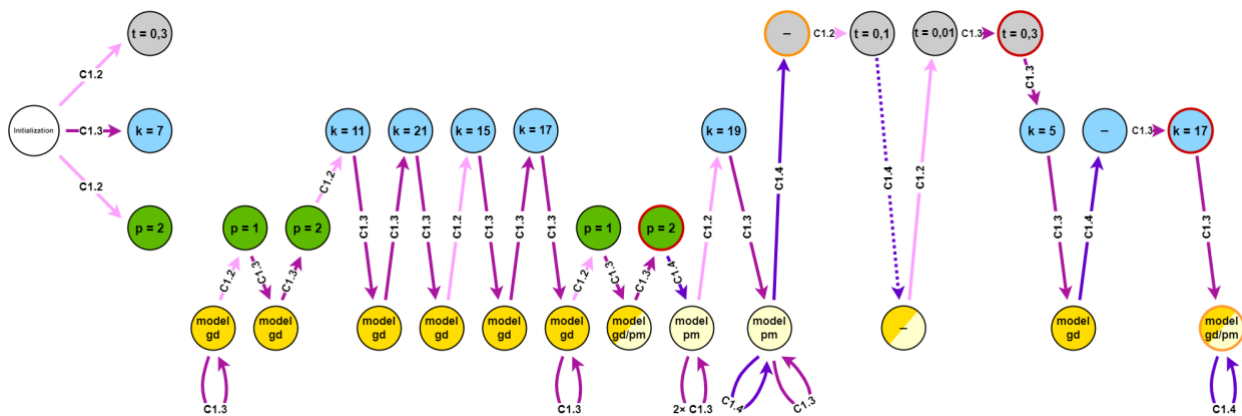


Figure 1: Processing graph of group 6

Each coded passage received both one code from the subcategories of main category C1 (*activity*) and one from the subcategories of main category C2 (*topic*). In the processing graph, each coded passage is represented by a node (representing the *topic*) and an edge oriented to the node (representing the *activity*). All subcategories are represented by unique colors. We start by looking at the nodes: The subcategory C2.1 (*testsize, t*) is represented by grey nodes, the blue nodes visualize the subcategory C2.2 (*number of neighbors, k*), the green nodes represent the subcategory C2.3

(*metric, p*), and the yellow nodes stand for the subcategory C2.4 (*model*). Subcategory *others* was assigned so rarely that it is not visualized. The labels of the nodes give additional information. For the subcategories where a parameter had to be chosen (*testsize, number of neighbors, metric*) the labels indicate the choice of the respective parameter. For the nodes regarding the *model* the labels indicate whether the students use performance measures (*pm*, light yellow) or graphical displays (*gd*, dark yellow) (see additional code C2.4*) when talking about their model. For all kinds of nodes, the label can be “-”, which means that the students’ process is not referring to a certain parameter or model, which is always and only the case for passages of subcategory C1.4 (*warrant expression*). For an additional structuring and a more compact overall picture, the nodes of the same subcategories are additionally arranged in rows.

Looking at the edges we find the following colors: Subcategory C1.1 (*decision, without content reason*) is displayed in light lilac (this subcategory doesn't appear the processing graph of group 6, but in the graph of group 3, see Figure 2), subcategory C1.2 (*decision, targeted trying*) in pink, subcategory C1.3 (*decision, justified*) in purple, and subcategory C1.4 (*warrant expression*) in dark purple. The additional label contains the same information, but the color supports the overview at first glance. Additional information, which does not result from the coding itself, but from the subsequent content analysis (Bata et al., 2021), lies behind in the dashed edges of category C1.3 (*decision, justified*) and C1.4 (*warrant expression*). The dashed edges mark passages where the used warrant is incomprehensible, incorrect, or pre-said by the design experiment leader. This additional visualization was integrated because otherwise a first wrong impression of the processing could be received.

Additional information is provided by the margins of the nodes: A red margin shows that a parameter has been finally set, which means that the value of the parameter is the one for the students’ final model. An orange margin shows a passage where the students talk about systematic parameter optimization (see additional code C1*).

Finally, the structure on the far left of a processing graph differs from the rest of the graph. The graph always begins with one white node labeled with “Initialization”, followed by three nodes regarding the parameters *testsize, number of neighbors, and metric*. This structure arises from the fact that all model building processes begin with the students initially selecting all three parameters to create an initial model. In the following, we call this first phase "initialization phase".

Results

The analysis of the model building processes with the processing graphs shows that in the twelve investigated groups not only twelve different models were build, but also twelve very individual processing graphs emerged. The processing graphs, shown here exemplarily for group 6 (Figure 1) and group 3 (Figure 2), differ for example with respect to the length and distribution of the category *activity*, visible by the number and color of the edges, and the order of the setting of the *parameters*, visible by the order of the nodes. We now describe five characteristic structures, which are, according to our definition, phases or properties occurring in the students' approaches which can be found in more than one of the observed groups. Each characteristic structure is first described with respect to its characteristics, and then we explain how it can be identified in the processing graph.

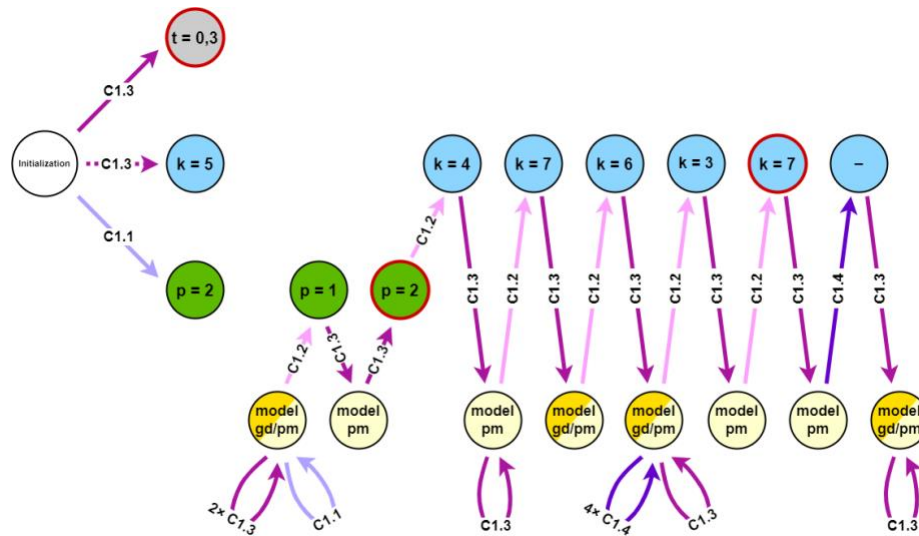


Figure 2: Processing graph of group 3

Finalization in the initialization phase

The first characteristic structure is, that some groups finalize one or two of the parameters *testsize* (C2.1) and *metric* (C2.2) directly during the initialization phase (see group 3, Figure 2). In this case, the associated *activity* is almost always the *justified decision* (C1.3) and it seems to result from the students' belief that there are clear correct choices for these two parameters. The finalization directly in the initialization phase can be identified by nodes with a red margin in the initialization phase, such as is group 3 (Figure 2) for the parameter *testsize*. group 6 (Figure 1) instead tried different values for all three parameters during the further process.

Justification increasing

The second characteristic structure is that, *activities* of the category C1.1 (*decision, without content reason*) and C1.2 (*decision, targeted trying*) frequently occur in the initialization, but when progressing, the parameter choices for these activities can increasingly be described by C1.3 (*decision, justified*). In this observed structure, students first seem to make sure that the code works and that an initial model is in place (C1.1 and C1.2) before they work on the task in more depth in terms of content (C1.3). This characteristic structure can be detected at one glance in the processing graphs by the increasing frequency of purple colored edges from left to right. It occurs in both groups shown.

Parameter selection phase with or without revisiting

The third characteristic structure is, that the parameters are set as far as possible one after the other and not in a mixed way. This characteristic structure can be recognized in the processing graphs by several nodes of the same color following each other only interrupted by yellow nodes (visible in both groups). We call this sequence of nodes for one parameter "parameter selection phase" and in

our sample at least one of these phases occurs in every group. The groups differ, however, in whether there is only one selection phase per parameter (group 3, due to finalization in the initialization phase, Figure 2) or whether after the (temporary) definition of a parameter another one is changed a second time (group 6, Figure 1), which could mean, that the students recognize that the selection of parameters influences each other. This characteristic structure might indicate students' understanding, since, for example, in the context of the given task for the variance of the model to remain constant, the number of neighbors must increase as the training size increases (i.e., the test size decreases).

Systematic trial and error versus discussion phases without verification by the model

The “parameter selection phases” described in the last section are primarily of two types. Either the students try out different parameter values with constant checking on the model or they discuss different parameter values without looking at the model. We call the first approach “systematic trial and error” and it can be identified in the processing graph by a permanent alternation between parameter nodes and model nodes (can be found in both groups). These phases differ partly with respect to the activities C1.2 (*decision, targeted trying*) or C1.3 (*decision, justified*), depending on the parameter (for example in the context of the parameter *metric* there are only two different values known that can be discussed, for the parameter *k* there are like 200) and whether the students use the performance measures (*pm*, light yellow) or graphical displays (*gd*, dark yellow) when looking at their model. The second approach, we call “discussion phase”, can be found for example in group 3 (Figure 1) for the parameter test size. Here, there is a sequence of grey nodes and the only yellow node in between is with an edge of category C1.4 (*warrant expression*), which means the model is addressed only theoretically.

Systematic parameter optimization

The last identified characteristic structure is “presence and position of conversations about systematic parameter optimization with a loop”. Systematic parameter optimization with a loop means, the students do not change the parameter again and again and look at the result (“systematic trial and error” as called in the previous section), but want to write a loop to try a number of parameters or parameter combinations at once and compare the results directly. Due to time constraints, this loop was not actually programmed in any of the groups, but the students talked about programming it. The conversations on this can be identified by the nodes with an orange margin in the processing graph (see group 6, Figure 1). They occur always at the beginning or end of a parameter selection phase and at the end of the task, which might be explained easily in terms of content.

Discussion

The coding of the model building processes with regard to the structure of the students' approaches had two goals: On the one hand, the passages in which an analysis of the used content is meaningful, should be identified. On the other hand, the students' structural approaches should be described and characterized. For the first goal, the straightforward coding with the category system was already sufficient; all passages of subcategories 1.3 and 1.4 could be analyzed subsequently in terms of content (Bata et al., 2021). For the description and characterization of the structure of the model building processes we have shown five characteristic structures, which can be observed in several

groups despite the many different ways of approaching the task or the model building process. As a part of the analysis the visualization of the coding with the processing graph proved to be especially helpful. The processing graph supported the identification of the characteristic structures, and all of them then can be found at one glance in the processing graph.

The characteristic structures presented have been developed from the analysis of all twelve groups and could be illustrated using groups 6 (Figure 1) and 3 (Figure 2). The results are subject to the limitation that all groups completed a very similar introductory course. The twelve different processing graphs and final models should not be viewed negatively as a lack of saturation, but rather as a result of the immense possible selections of parameters and possible approaches of model building. Nevertheless, it cannot and should not be excluded that there may be further characteristic structures, especially in different model building processes.

The goal of the overall design research project is to use a white-box approach to introduce students to the topic of machine learning and to find out how the developed material can support this approach (Bata et al., 2022). The aim of the analysis of the videos in the context of this goal was to start from an empirical perspective: The students' approaches in using the material are observed and described in order to determine, for example, whether different structural and content-related approaches are conditional and which of them appear to be conducive to the task and the subsequent use of the model. For this aim, the structural description as given in this paper is an important step, which will now be followed by a systematic and comparative view of all groups, as well as a mapping between the content-related results, such as model warrants or parameter warrants (Bata et al., 2023), and the individual characteristic structures. In this way, it may be possible to identify desirable approaches for the students in terms of structure and content and to support them, for example, by restructuring the task.

References

- Bata, K., Schmitz A. & Eichler A. (2023). Einblicke in die Bearbeitung einer Aufgabe zur Erstellung von Modellen mit Maschinellen Lernverfahren [Insights into the work on a task about building models with machine learning methods]. In Goethe-Universität Frankfurt (Hrsg). *Beiträge zum Mathematikunterricht 2022. 56. Jahrestagung der Gesellschaft für Didaktik der Mathematik*. WTM. <https://doi.org/10.37626/GA9783959872089.0>
- Bata, K., Schmitz, A., & Eichler, A. (2022). Insights into the design of an introductory course for data science and machine learning for engineering students. In G. Bolondi, F. Ferretti & J. Hodgen (Eds.), *Proceedings of the 12th Congress of the European Society for Research in Mathematics Education (CERME12)*. Free University of Bozen-Bolzano and ERME.
- Bata, K., Eichler, A., & Schmitz, A. (2021). How Engineering Students Argue in an Introductory Course in Data Science. In R. Helenius & E. Falck (Eds.), *Proceedings of the IASE 2021 Satellite Conference*. International Association for Statistical Education.
- Biehler, R., & Schulte, C. (2018). Perspectives for an interdisciplinary data science curriculum at German secondary schools. In Biehler, R., Budde, L., Frischemeier, D., Heinemann, B.,

- Podworny, S., Schulte, C., & Wassong, T. (Eds.), *Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts*. Paderborn: Universitätsbibliothek Paderborn.
- Bilstrup, K.K., Kaspersen, M.H., Assent, I., Enni, S., & Petersen, M.G. (2022). From Demo to Design in Teaching Machine Learning. *2022 ACM Conference on Fairness, Accountability, and Transparency*.
- Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, 16(1), 44-49.
- Fiebrink, R. (2019). Machine Learning Education for Artists, Musicians, and Other Creative Practitioners. *ACM Transactions on Computing Education*, 19(4), 31:1-31:32. <https://doi.org/10.1145/3294008>
- Gravemeijer, K. P. E., & Cobb, P. (2006). Design research from a learning design perspective. In Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.). *Educational Design Research* (pp. 45-85). Taylor and Francis Ltd.
- Huppenkothen, D. & Eadie, G. (2020). Teaching the Foundations of Machine Learning with Candy. *Proceedings of Machine Learning Research*, 141, 29-35.
- Heuer H., Jarke J. & Breiter A. (2021). Machine learning in tutorials – Universal applicability, underinformed application, and other misconceptions. *Big Data & Society*, 8(1). <https://doi.org/10.1177%2F205395172111017593>
- Kinnaird, K. M. (2020). Teaching Computational Machine Learning (without Statistics). *Proceedings of Machine Learning Research*, 141, 17-22.
- Lavesson, N. (2010). Learning Machine Learning: A Case Study. *IEEE Transactions on Education*, 53(4), 672–676. <https://doi.org/10.1109/TE.2009.2038992>
- Silva Marques, L., Gresse von Wangenheim, C. & Hauck, J. (2020). Teaching Machine Learning in School: A Systematic Mapping of the State of the Art. *Informatics in Education*. 19. 283-321. <https://doi.org/10.15388/infedu.2020.14>
- Steinbach, P., Seibold, H., & Guhr, O. (2020). Teaching Machine Learning in 2020. *Proceedings of Machine Learning Research*, 141, 1–6.
- Sulmont, E., Patitsas, E., & Cooperstock, J. E. (2019). What Is Hard about Teaching Machine Learning to Non-Majors? Insights from Classifying Instructors' Learning Goals. *ACM Transactions on Computing Education*, 19(4), 1–16. <http://dx.doi.org/10.1145/3336124>
- Toulmin, S. (2003). *The uses of argument* (2nd ed.). Cambridge, England: Cambridge University Press.
- Touretzky, D.S., Gardner-McCune, C., Martin, F.G., & Sehorn, D.W. (2019). Envisioning AI for K-12: What Should Every Child Know about AI? *AAAI Conference on Artificial Intelligence*, Honolulu, HI, USA. <https://doi.org/10.1609/aaai.v33i01.33019795>