# HOW ENGINEERING STUDENTS ARGUE IN AN INTRODUCTORY COURSE IN DATA SCIENCE

Katharina Bata, Andreas Eichler and Angela Schmitz
TH Köln – University of Applied Science, Germany
Universität Kassel, Germany
katharina.bata@th-koeln.de

*Model building and validation are at the core of machine learning and a subfield of data science. In this paper, the Toulmin model is used to structure students' approaches and analyze students' argumentation when building a model. A qualitative analysis of passages from the underlying design experiments with undergraduate engineering students shows different approaches and visual, contextual and mathematical or statistical elements that students use within their argumentation.*

## INTRODUCTION

Due to the increasing popularity and relevance of data science in industry and research, education in the field of data science is gaining importance (Engel, 2017; Grillenberg & Romeike, 2018). Especially for engineering students the focus is on practical applications like data analysis and machine learning, which are part of a large number of data science curricula (Grillenberg & Romeike, 2018), but still missing in many engineering curricula (Heidling et al., 2019).

As the teaching of machine learning is a topic that has been little explored (Steinbach et al., 2020), there are many open questions regarding a syllabus, teaching methods, and learning processes. Our research focuses on an attempt to teach an introduction of data science and, particularly, machine learning for engineering students in a few lectures. In this paper, we present one part of our project, that is, students' argumentation during the introductory course.

## THEORETICAL BACKGROUND AND RESEARCH QUESTION

### *The teaching of machine learning*

In recent years, the call to systematically explore the teaching of machine learning (ML) has been expressed in various contexts (Ko, 2017; Fiebrink, 2019; Steinbach et al., 2020). In addition to technical courses for computer science or mathematics students, there are many different best-practice examples for teaching ML to target groups with different backgrounds. The approaches differ specifically with respect to the mathematical and programming details. For example, these details are either left out in hands-on concepts (Fiebrink, 2019), first introduced using playful approaches (Huppenkothen & Eadie, 2020), or prepared via heuristics and pseudocode (Kinnaird, 2020). Beyond accompanying research on approaches such as those mentioned above, there are few explicit methods and empirical studies of how learning occurs using these methods (Steinbach et al., 2020).

One learning objective that is widely agreed upon is model validation (e.g., Fiebrink, 2019; Steinbach et al., 2020; Kinnaird, 2020; Huppenkothen & Eadie, 2020). Knowledge of ML model validation is not only helpful in assessing the expressiveness and usability of ML models, but also in building valid models. However, validation seems to be a particular hurdle for students because statistic tests and performance measures are used (Lavesson, 2010). From the perspective of many instructors a strong computer science and math background is crucial for validation (Sulmont et al., 2019).

### *Argumentation*

It has been empirically shown that there is a connection between students' argumentation and learning success, the understanding of subject-specific concepts, and individual knowledge construction (Budke & Meyer, 2015). A model for analyzing students' argumentation is provided in form of the so called functional argumentation analysis based on Toulmin's model of argumentation (Toulmin, 2003; Kopperschmidt, 1989).

In Toulmin's model (see Figure 1), argumentation can be represented in a structure consisting of *data*, *conclusion*, *warrant*, and *backing* (Toulmin, 2003; note: Another component not used here is *rebuttal*). *Data* is the unquestionable fact as the starting point of an argumentation with which the *conclusion* is justified, in shortest application "conclusion, because data". The *warrant* is a general

In: R Helenius, E Falck (Eds.), *Statistics Education in the Era of Data Science*
*Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),*
Aug-Sept 2021, Online conference.  iase-web.org/Conference_Proceedings.php

statement which serves as a bridge from *data* to *conclusion* and can be supported by the specification of a *backing*.

In a German oriented part of educational sciences and especially in mathematics education, the functional argumentation analysis has been widely used to reconstruct and analyze students' argumentation processes (Budke et al., 2015; Fetzer, 2011).

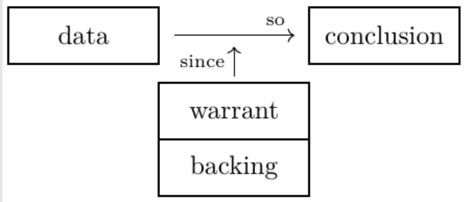


Figure 1. Toulmin's model

*Research questions*

With respect to the described hurdle of building and validating ML models, arguing has a dual function. On the one hand, arguing at this point is itself a learning objective, and on the other hand, it can provide information about students' learning processes. Because of this dual role, the following two research questions will examine the students' argumentation in more detail:

1. How do students proceed when building a first ML model?
2. What elements do students use in argumentation processes when building ML models?

## METHOD

*The overall research project*

The research questions are investigated in an introductory course the prototype of which is developed within a project following the methodology of design-based research (Gravemeijer & Cobb, 2006). The prototype emerges from a topic-specific analysis of selected ML methods and different design elements using the "four-level approach for specifying and structuring mathematical learning content" (Hußmann & Prediger, 2016) and has been tested and improved over two design cycles (further cycles planned). In this paper, we refer to results from the first cycle, in which design experiments were conducted in a laboratory setting with seven undergraduate students from different engineering programs. The participating students had no prior experience with data science and were above average in mathematics but brought below average programming experience compared to an average engineering student of the target group of our introductory course. They were guided through the material by the lecturer via an online conference tool in groups of two or three and accompanied during the processing of the tasks. Each session was video-recorded; in addition, the group work was documented by means of written products.

*The observed task*

The analysis of the students' approach and argumentation is based on a task within the developed introductory course. In the task, 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 students are asked to fit a model with the k-nearest-neighbor method concerning the dataset in order to classify a new dataset using their model afterwards. The Jupyter Notebook prepared for the task contains code cells that define the basic framework for model building with k-nearest-neighbor and all parameters are initially empty. It also contains code cells that can be used to show a choice of performance measures (cf. figure 2), as well as one graphic display of the model once with the training data and once with the test data (cf. figure 4). The students work together on a split screen, talking to each other about the model and how to set the parameters.

*The qualitative content analysis and functional argumentation analysis*

In a first step, the videos were analyzed based on the research question “How do students proceed when building a first ML model?" using a content-structuring qualitative content analysis (Kuckartz, 2014). Two goals were pursued with the analysis: On the one hand, sections of the task processing relevant to the research question had to be separated from irrelevant sections (e.g. thoughts about syntax). On the other hand, the analysis was intended to provide an overview of the entire process of model building in order to be able to examine it with regard to different procedures. In the context of the analysis, a deductive-inductive category system consisting of main categories and subcategories (Kuckartz, 2014) was developed, which was used to code the video of the processing of the described task (about 30 minutes of video material per group). The category system was validated by means of consensual coding (Kuckartz, 2014), and finally by determining the intercoder reliability. The calculated intercoder reliabilities of the three main categories range from 0.76 to 0.86. Taking into account the number of subcategories, this corresponds to a Cohens kappa between 0.68 and 0.8.

In the second step, the students' task processing was analyzed in more detail with regard to the research question "What elements do students use in argumentation processes when building ML models?" The passages that were identified as relevant in the qualitative analysis and that contained argumentative elements in Toulmin's sense were transcribed and analyzed using the functional argumentation analysis. The analysis is not limited to explicit statements; the reconstruction of implicit components of an argumentation process is also possible (Fetzer, 2011). A judgement about the correctness and resilience of the warrants used by the students is not made.

The non-spoken interactions, like students' gestures, which might also be interesting because they could influence their conversations, were not considered in these two steps, since the students were sitting at their own computers, spatially separated, and only their faces were visible in the webcam.

## RESULTS

*The category system for research question 1*

A result of the analysis of how students proceed when building a first ML model is the category system shown in Table 1. It allows to identify the students' approach based on the questions "How?" (*type of approach*), "Related to what?" (*topic*), and "When?" (*time*).

The main category *type of approach* marks the sections in the video that are essential for the building process and distinguishes four inductively developed subcategories. The subcategories *decision (trying)* and *decision* (*preference unfounded*) mark for example affective decisions. Sections coded with the subcategories *decision (justified)* and *pure warrant* contain argumentative elements. The main category *topic* codes what the sections coded in *type of approach* refer to in terms of content. If a decision was coded, the main category *time* differentiates whether the decision takes place before fitting or on the basis of the already fitted model.

Table 1. Overview of the three main categories

| Main category | Subcategory |
|---|---|
| **type of approach** | decision (trying)<br>decision (unfounded preference)<br>decision (justified)<br>pure warrant |
| **topic** | model<br>testsize<br>number of neighbors (k)<br>metric (p)<br>other |
| **time** | before fit<br>after fit |

*Insights into the argumentation processes*

The subcategories *decision (justified)* and *pure warrant* mark the sections that are relevant with respect to the research question "What elements do students use in argumentation processes when building ML models?" Different types of argumentation occurred during the functional argumentation analysis. The following illustration focuses on two aspects in more detail:

1. Students' argumentation based to a large extent on the graphic displays or on the performance measures. For the sample studied, it can be observed that the groups focus on one of the two approaches, i.e., they argue either based on the graphic displays or based on the performance measures.
2. There is argumentation in which students use mathematical and statistical elements to support arising warrants.

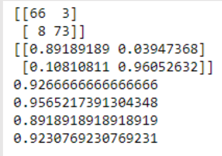
*Aspect 1: Graphic displays vs. performance measures*

For the analysis, examples from two groups are shown. Group HWB, consisting of students with the pseudonyms Hanna, Wael and Bastian, mainly uses arguments based on performance measures. Group NS, consisting of the students with the pseudonyms Nele and Sandra, mainly uses arguments based on the graphic displays. The following passage shows an often observed approach of the students: Exploratory fixing of a parameter (coded: *type of approach* as *decision (trying)* with *time* as *before fit* and *topic* as *metric*):

Example 1: Hanna, Wael and Bastian (argumentation using performance measures)

1.1 Wael: Should we just take a different metric and look at the numbers?
1.2 Bastian: Yeah, actually we can do that.

*(W. screenshots the values, fits the model with p=1 (absolute value metric) and prints out the performance measures, see Figure 2)*



```
[[66  3]
 [ 8 73]]
[[0.89189189 0.03947368]
 [0.10810811 0.96052632]]
0.9266666666666666
0.9565217391304348
0.8918918918918919
0.9230769230769231
```

Figure 2. Confusion-matrices, precision, recall and f1-score of both models

Then, the fitted model is used to assess whether the explorative fixed parameter is good or bad (coded: *type of approach* as *decision (justified)* with *time* as *after fit* and *topic* as *model* and as *metric*):

1.3 Wael: Yes, well, it's worse, you can see that instantly, right?
1.4 Bastian: Yes (…) Yes, okay, that's / that’s also a good finding.
1.5 Wael: <u>All</u> values are worse, if I see it correctly.
1.6 Hanna: But really <u>all</u> of them, right?
1.7 Wael: Yes (laughing).

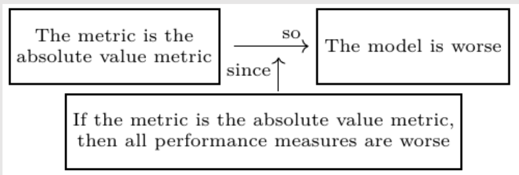



Figure 3. Toulmin's model of argumentation

With the functional argumentation analysis the model of figure 3 results. Performance measures are used for argumentation in the elaborated warrant.

In the example of group NS the procedure is analogous (concerning a different parameter). But, as can be seen in the functional argumentation analysis, here the warrant is based on the graphic display (see figure 5):

Example 2: Nele and Sandra (argumentation using graphic display)

2.1 Nele: But we can see now what happens, if with the right metric, that we actually wanted to have, we go back to 20.
2.2 Sandra: Mhm.

*(N. fits the model with k=20 and prints out the graphic display, see Figure 4.)*

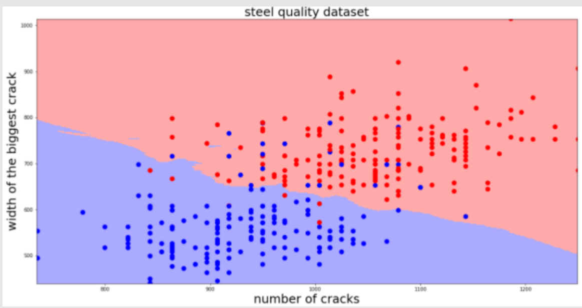



Figure 4. Graphic display of the model with training data

2.3 Nele: Umm, no, 30 is better.
2.4 Sandra: Why do you think so?
2.5 Nele: Yeah there are curves in here now, and there are blue spots and so on, and there's a funny cookie cutter thing here.

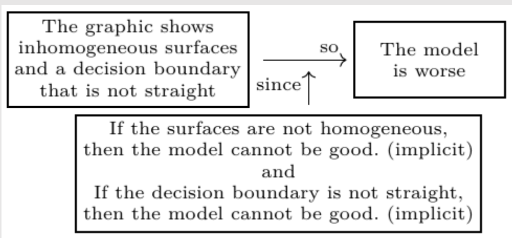



Figure 5. Toulmin's model of argumentation

The used warrants differ, besides the difference regarding the use of the graphic displays and the performance measures, in that the warrants in example 2 are only implicitly recognizable but can be reconstructed. In both examples, the conclusion is that the newly fitted model is worse than the original model. The obvious conclusion, that the parameter is set back to the original value, follows subsequently in both groups.

Exemplarily, it was shown in both groups that one approach of the students in the building of ML models is a first explorative setting of parameters followed by a reasoned assessment of the resulting model. The argumentation in the reasoned assessment is based, among other things, on the graphic displays and the performance measures. As a further approach, it was observed, for example, that students define parameters in advance on the basis of formulated warrants. The example under aspect 2 shows such an approach.

*Aspect 2: Mathematical and statistical elements*

The following passage from group HWB shows the backing of an expressed warrant with a reference to a phenomenon of statistics (coded: *type of approach* as *pure warrant* with *topic* as *testsize*): Example 3: Hanna, Wael and Bastian (argumentation with backing from statistics)

3.1 Wael: The more data you take, the better the model will be, I would say, simply. But the more umm / the more data you take, the less data you have to validate it afterwards. So the more training data, I mean now.
3.2 Bastian: The question is whether the model gets better and better when you take more training data, then you also get into such an overfitting, right? So under some circumstances.
3.3 Wael: No, that has nothing to do with overfitting, does it? It's more like the more umm / I mean, you have this small-n-problem in the statistics, if you have too few things, then you have individual outliers, which are maybe very rare overall, but if they are just in there now and thereby your / they ruin your thing, your statistics, and that also applies to the model, right?
3.4 Bastian: Yes.

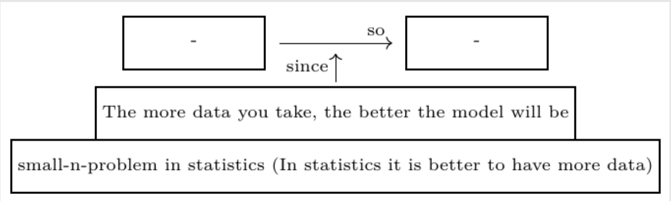


Figure 6. Toulmin's model of argumentation

Wael starts by stating a general warrant about the influence of training data on the model. Bastian doubts this warrant by referring to overfitting, whereupon Wael supports his warrant with a backing by a comparison from statistics (see figure 6).

This section shows a warrant expression that is used by the students in the further course to set a parameter in a justified way before fitting, unlike in examples 1 and 2.

## DISCUSSION

For research question 1, a category system was developed, which can be used to describe the approaches observed so far by the students when building a first ML model in a structured way. The examples show exemplarily that both trying and justifying approaches are used and that argumentation is given both before and after fitting as well as warrants without direct decisions are discussed. For research question 2, graphic displays, performance measures, and comparisons from statistics were used to exemplify elements that students use in argumentation processes. Thus, visual as well as contextual and mathematical or statistical elements seem to play a role in the argumentation.

Methodological limitations of this result follow from the fact that all students participated in the same introductory course. Furthermore, all students participating in the first design cycle were above average mathematically but brought below average programming experience. Regarding both aspects, it is expected that in the following cycles there will be a wider spectrum of approaches and aspects regarding the elements used in the argumentation.

It should also be noted that only examples from a small and early section of the entire model building process are shown. Other important aspects, such as students' discussions about the influence of parameters on the expressiveness and usability of a ML model, occur later in the learning path and can also be observed in later excerpts of the videos.

Nevertheless, the results presented exemplify that even with a short introductory course students can be encouraged to use visual, contextual and mathematical or statistical means when discussing issues regarding data science, especially ML. But even though argumentation is related to the learning process (Budke & Meyer, 2015), with the analysis so far it remains unclear how deep the students' understanding of the used warrants is. An investigation of students' learning paths, as planned in the underlying design-research project, will provide further insights regarding this question and promises to reveal more about the reasons why students perceive model validation as a hurdle (Lavesson, 2010). Overall, our research offers starting points for the development of learning environments for data science and machine learning that go beyond a best practice experience.

REFERENCES

Budke, A., Kuckuck, M., Meyer, M., Schäbitz, F., Schlüter, K., & Weiss, G. (2015). *Fachlich argumentieren lernen. Didaktische Forschungen zur Argumentation in den Unterrichtsfächern*. Münster, Germany: Waxmann.

Budke, A. & Meyer, M. (2015). Fachlich argumentieren lernen – Die Bedeutung der Argumentation in den unterschiedlichen Schulfächern. In Budke, A., Kuckuck, M., Meyer, M., Schäbitz, F., Schlüter, K., & Weiss, G. *Fachlich argumentieren lernen. Didaktische Forschungen zur Argumentation in den Unterrichtsfächern* (pp. 9-28). Münster, Germany: Waxmann.

Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal, 16*(1), 44-49.

Fetzer, M. (2011). Wie argumentieren Grundschulkinder im Mathemathematikunterricht? Eine argumentationstheoretische Perspektive. *Journal für Mathematik-Didaktik*, *32*, 27-51.

Fiebrink, R. (2019). Machine Learning Education for Artists, Musicians, and Other Creative Practitioners. *ACM Transactions on Computing Education, 19*(4), 31:1-31:32.

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). London, England: Taylor and Francis Ltd.

Grillenberg, A. & Romeike, R. (2018). Ermittlung der informatischen Inhalte durch Analyse von Studienangeboten. *Commentarii informaticae didacticae*, *10*(1), 119-134.

Heidling, E., Meil, P., Neumer, J., Porschen-Hueck, S., Schmierl, K., Sopp, P., & Wagner, A. (2019). *IMPULS-Studie Ingenieurinnen und Ingenieure für Industrie 4.0.* http://www.impuls-stiftung.de/documents/3581372/4875823/Ingenieurinnen0und0Ingenieure0f%C3%BCr0Industrie04.0IMPULS0Studie0Januar02019.pdf/9a5672ea-6fe3-a379-2a70-a8536fbfe73e.

Huppenkothen, D. & Eadie, G. (2020). Teaching the Foundations of Machine Learning with Candy. *Proceedings of Machine Learning Research, 141*, 29-35.

Hußmann, S. & Prediger, S. (2016). Specifying and Structuring Mathematical Topics. *Journal für Mathematik-Didaktik, 37*(1), 33-67.

Kinnaird, K. M. (2020). Teaching Computational Machine Learning (without Statistics). *Proceedings of Machine Learning Research, 141,* 17-22.

Kopperschmidt, J. (1989). *Methodik der Argumentationsanalyse.* Stuttgart-Bad Cannstadt, Germany: Frommann-Holzboog Verlag.

Ko, A. J. (2017). *We need to learn how to teach machine learning.* https://link.medium.com/wmzrUMMj3bb.

Kuckartz, U. (2014). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (2nd ed.). Weinheim, Germany: Beltz Verlagsgruppe.

Lavesson, N. (2010). Learning Machine Learning: A Case Study. *IEEE Transactions on Education, 53*, 672-676.

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.

Toulmin, S. (2003). *The uses of argument* (2nd ed.). Cambridge, England: Cambridge University Press.