



HAL
open science

AI education as a starting point for interdisciplinary STEM projects

Sarah Schönbrodt, Katja Hoeffler, Martin Frank

► **To cite this version:**

Sarah Schönbrodt, Katja Hoeffler, Martin Frank. AI education as a starting point for interdisciplinary STEM projects. 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-04420534

HAL Id: hal-04420534

<https://hal.science/hal-04420534>

Submitted on 26 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.

AI education as a starting point for interdisciplinary STEM projects

Sarah Schönbrodt¹, Katja Hoeffler² and Martin Frank³

¹Paris Lodron University of Salzburg, Austria; sarah.schoenbrodt@gmail.com

²Lessing-Gymnasium, Karlsruhe, Germany

³Karlsruhe Institute of Technology, Germany

Artificial intelligence (AI) is a key technology nowadays and of enormous relevance for our society. A central characteristic of applications and methods from the field of AI is that they are highly interdisciplinary. Both mathematics, computer science, and other STEM as well as non-STEM disciplines are crucial depending on the application at hand. In addition, ethical considerations play an important role. Thus, together with the increasing demand to bring AI education into schools, comes the opportunity to design and implement interdisciplinary projects. In this paper, a developed digital learning module on the problem of human activity recognition on smartphones is used to show how interdisciplinary AI projects can be designed. We use a context-integration approach that focuses on mathematics embedded in mathematical modeling activities, while incorporating knowledge from computer science and physics.

Keywords: Artificial Intelligence, Data Science, mathematical modeling, STEM education, Jupyter Notebooks.

Introduction

Methods from the field of artificial intelligence (AI) underlie numerous applications in our everyday life: from the classification of human activities in fitness apps to face recognition systems for unlocking smartphones and large-scale language models as used in ChatGPT. The enormous and yet increasing relevance of such applications for our society results in the demand for a stronger integration of AI education into school.

A central characteristic of AI methods and its applications is, that they are highly interdisciplinary. Mathematics and computer science, but also knowledge from the field for which the application is to be developed, play a key role. For example, biological knowledge is important in the development of AI applications in the medical field and applications from the field of natural language processing offer opportunities to take up AI as a topic in linguistics. Hence, fruitful connections to various subjects, STEM, and non-STEM, can be identified. This leads us to the conclusion that AI education cannot or should not take place exclusively in computer science classes, but that it can certainly be integrated in further subjects (especially mathematics!) and in interdisciplinary STEM projects.

Interdisciplinary mathematical modeling and AI education

This paper focuses on the opportunities to combine knowledge from mathematics, computer science and physics in an interdisciplinary way when designing mathematical modeling activities on AI-applications. In short, the term mathematical modeling is used here to describe the process of solving real-world problems with the help of mathematics by translating the real-world into mathematics and

vice versa. Mathematical modeling is thus by definition interdisciplinary and requires knowledge outside of mathematics (English, 2022, p. 164).

In this paper, the problem of human activity recognition on smartphones is considered, for which we developed a digital learning module (Hoeffler, 2022, pp. 37–52). The learning module is mainly based on interactive, educationally designed Jupyter Notebooks¹ and enables upper secondary students to develop an AI method based on real sensor data. Jupyter Notebooks are, in simple terms, digital documents that can be used to write text and run code (e.g., to visualize large datasets) and can be edited interactively by students. They are used in numerous sectors of industry and research. Thus, the digital tool offers an authentic insight into current problem-solving strategies and technical implementations in applied mathematics or more generally in STEM fields. The developed Jupyter Notebooks are based on the programming language Python. For direct use in the classroom the Notebooks are made available to students and teachers via a browser-based platform. Thus, the material can be edited without installing any software (Schönbrodt et al., 2022b, p. 159).

In the following we describe the design principles of the digital learning material and the course of the learning module. By doing so we emphasize that and to what extent the up-to-date topic of AI is suitable for the design of interdisciplinary mathematical modeling projects.

Goals and design principles of the digital learning material

The learning module is intended to build up a basic understanding of methods from the field of AI, more precisely for its branch machine learning. The students should get to know and understand the principle of (supervised) machine learning using the example of an AI algorithm. In particular, it should become clear that mathematics is crucial for the development of AI methods.

The learning module should allow students to experience the relevance of mathematical modeling for their everyday lives by solving a real-world problem. In addition, it should allow students to gain a deeper understanding of mathematical modeling. For this purpose, the learning module enables the complete and repeated run through the mathematical modeling cycle. In addition, a four-step modeling cycle according to Blum (1985) is used as a strategic, metacognitive element to support the learners' reflection on the entire modeling process. Therefore, the performed modeling steps are reflected repeatedly within short discussions sessions.

The learning module should allow students to experience that the process of mathematical modeling as well as the development of AI methods is highly interdisciplinary. For this purpose, the interaction of the disciplines of mathematics, computer science and physics to solve a real-world problem is pointed out. Interdisciplinary aspects are not excluded but integrated and interconnected - whenever it is reasonable and authentic within the mathematical modeling process. Thus, a context integration model is chosen for designing STEM education around AI: When solving a real-world problem, students should especially grasp the mathematical basics and reflect on their own modeling process.

¹ For further information see <https://jupyter.org>, last accessed: 10 February 2023

For the sake of authenticity and relevance, content from other disciplines is integrated (Roehrig et al., 2012, p. 35).

The learning module should account for the heterogeneity of the learning group and enable the students to work on the development of the AI model in a self-determined manner. To this end, additional more advanced tasks and scaffolded tips have been integrated into the digital material. In addition, the learning material is designed in such a way that no programming knowledge is required. Instead, a fill-in-the-gap approach was implemented. The students simply need to replace a placeholder within the code with the relevant solution to the task (a number, formula, or function). They receive automated task-specific feedback on their input and, if necessary, a tip on how to correct their solution (Schönbrodt et al., 2022b, p. 159). Optionally, interested students can write code themselves. In this way, both students with and without programming skills are challenged according to their knowledge.

Insights into the course of the learning module

In this section the problem and the main mathematical modeling steps of the learning module are described. At the same time, connections to disciplines of computer science and physics are highlighted.

The problem: human activity recognition on smartphones

Mobile devices, such as smartphones, have become constant companions in our everyday lives. In addition, interest in evaluating people's habits and daily routines has increased in recent years. This includes, among other things, the analysis of human activities. Due to the enormous advancement of sensors in smartphones, these can be used to recognize human activities (Hoeffler, 2022, p. 1).

The aim of the learning module is to classify human activities based on sensor data recorded with the smartphone, more precisely with the app phyphox². Phyphox is an app that was especially developed for the use in physics teaching (Staack et al., 2018, p. 1).

Unit 1: Understanding the data

Acceleration data in all three spatial dimensions serve as the starting point for the development of the classification model. These were recorded over a longer period of time (several minutes) using the phyphox accelerometer for five types of movement (sitting, standing, walking, running, climbing stairs). During all measurements, the smartphone was in the front pocket of the person's pants. For the recorded data, it is known to which activity type (i.e., class) the data belongs. Thus, each data point initially consists of the corresponding activity class (ActivityID), the time in seconds, the accelerations in all three spatial dimensions ax , ay and az as well as the length of the acceleration vector $\vec{a} = (ax, ay, az)^T$. The corresponding dataset is shown in figure 1.

² Phyphox is an app that allows to perform physical experiments with the help of smartphone sensors, <https://phyphox.org>

In understanding the given data physical knowledge is crucial, especially regarding acceleration and the influence of gravitational acceleration (e.g., when sitting and standing, acceleration is not zero, due to the Earth's gravity).

| ActivityID | Time (s) | ax (m/s ²) | ay (m/s ²) | az (m/s ²) | Vector Length (m/s ²) |
|------------|----------|------------------------|------------------------|------------------------|-----------------------------------|
| 1 | 0.028140 | 5.480853 | -4.508624 | -6.869365 | 9.877025 |
| 1 | 0.053292 | 5.488187 | -4.504134 | -6.870263 | 9.879673 |
| 1 | 0.078443 | 5.494324 | -4.496799 | -6.872658 | 9.881408 |
| 1 | 0.103595 | 5.492229 | -4.508475 | -6.871311 | 9.884626 |
| 1 | 0.128747 | 5.472320 | -4.499044 | -6.856192 | 9.858756 |
| ... | ... | ... | ... | ... | ... |

Figure 1: Excerpt of the dataset from the digital learning material. The ActivityID represents the type of movement (1 = sitting, 2 = standing, 3 = walking, 4 = running, 5 = climbing stairs)

Unit 2: Preprocessing the data

The recorded data is first divided into smaller time windows (3 sec.). Certain statistical measures are then discussed and determined for each time window: the maximum, minimum and average length of the acceleration vector. Thus, for each time window, a 3-dimensional data point is available (see Figure 2). This dataset serves as a starting point for the development of a mathematical model for the classification of activities. The decision to reduce the data initially to only three features was made for didactical reasons: the data can be graphically presented and explored (see Figure 3).

| WindowID | ActivityID | Mean Vector Length (m/s ²) | Max Vector Length (m/s ²) | Min Vector Length (m/s ²) |
|----------|------------|--|---------------------------------------|---------------------------------------|
| 1.0 | 1.0 | 9.879301 | 9.942203 | 9.822518 |
| 2.0 | 1.0 | 9.879365 | 9.986057 | 9.791405 |
| 3.0 | 1.0 | 9.880404 | 9.962945 | 9.791129 |
| 4.0 | 1.0 | 9.879834 | 10.023475 | 9.765301 |
| 5.0 | 1.0 | 9.881129 | 9.916207 | 9.845664 |
| ... | ... | ... | ... | ... |

Figure 2: Excerpt of the dataset. For each time window the mean, the maximum and the minimum length of the acceleration vector has been computed

When developing AI algorithms, especially within its subarea machine learning, the data is usually divided into so-called *learning* or *training data* and *test data*. Based on the training data, a model is developed ("learned"), which should predict as accurately as possible to which activity type new, unknown data belongs. This explains the term *machine learning*. The test data, for which the class assignment is also known but which was not used in the model development phase, is used to validate the model.

Unit 3: Developing a mathematical model

In practice, various AI methods are used to classify human activities. Within the learning module we used the so-called k-Nearest Neighbor (kNN) algorithm. The kNN algorithm is accessible with high-school mathematical knowledge, and it also leads to good results in real-world applications (Hazzan & Mike, 2022, p. 19; Mohsen et al., 2022, p. 305). The name of the algorithm already suggests its approach: Using an appropriate mathematical measure, it is defined what it means for two data points

to be similar (so-called neighbors). Then, for a new data point to be classified, the k most similar data points, i.e., the k nearest neighbors, are determined.

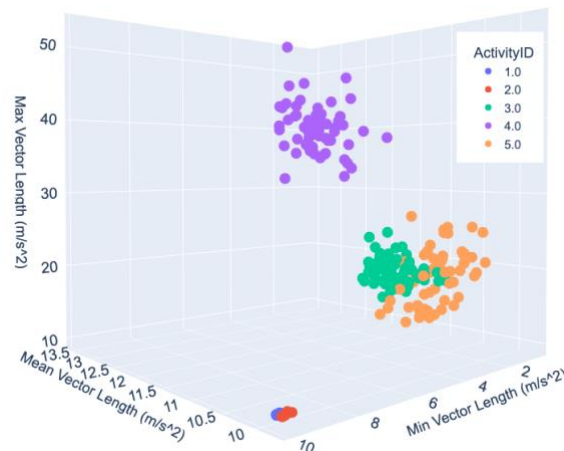


Figure 3: For different time windows of the given acceleration data, the maximum, the minimum and the mean length of the acceleration vector were calculated

When choosing an appropriate distance function to model similarity, students can first creatively follow their own ideas. In addition, the following distance functions that are commonly used in research are introduced in the learning material,

- the Euclidean distance

$$d_E(\vec{a}_1, \vec{a}_2) = \|\vec{a}_1 - \vec{a}_2\|_2,$$

- the Manhattan distance

$$d_M(\vec{a}_1, \vec{a}_2) = \|\vec{a}_1 - \vec{a}_2\|_1,$$

- and the cosine distance³

$$d_C(\vec{a}_1, \vec{a}_2) = 1 - \frac{\vec{a}_1 \cdot \vec{a}_2}{\|\vec{a}_1\|_2 \cdot \|\vec{a}_2\|_2}.$$

Since 3-dimensional data have been used so far, these distance functions can be discussed based on school mathematical knowledge (vectors, length of a vector, dot product) and can also be visualized in a graphical way.

Now, to classify a new data point, the k most similar data points are determined (the k nearest neighbors). By majority vote, the new data point is then assigned to the class that had the most data points among the k nearest neighbors. This idea is illustrated in Figure 4 based on the Euclidean distance. At this point the learning material contains different versions: a) the students simply apply the pre-implemented algorithm, b) they implement it partially based on a fill-in-the-gap code, or c) they implement it on their own. In cases b) and c), central control structures (e.g., for-loops) are used.

³ Note that the cosine distance is not a true distance function as it does not hold the triangle inequality.

Unit 4–6: Evaluating the classification model

Finally, the developed classification model is validated on test data. Regarding the evaluation of the model and the classification results, different quality measures can be discussed with the students. In the learning module, accuracy, error rate and precision are introduced and determined. In addition, the results are presented and evaluated in a confusion matrix.

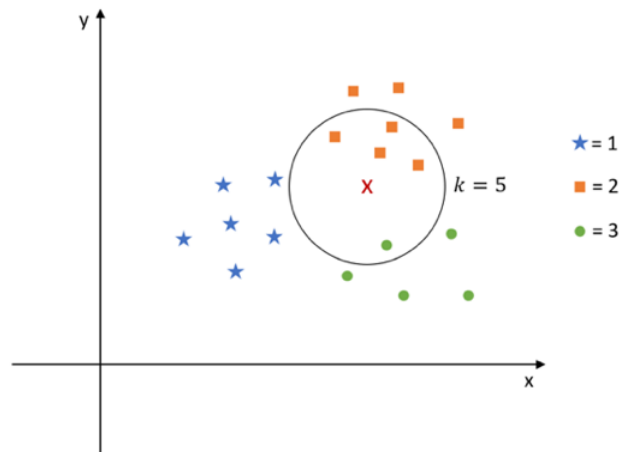


Figure 4: Basic idea of the kNN algorithm

Unit 5–6: Model improvements

Until now, a 3-dimensional dataset was used, which contained only a few statistical measures. To improve the classification results the students can extend the dataset by additional features (i.e., the minimum, the maximum and the arithmetic mean of each of the three acceleration directions, and the median, the lower as well as the upper quartile of the length of the acceleration vector). Furthermore, the students can apply different distance functions and vary the value of the parameter k to enhance the classification results.

Unit 7: Critical evaluation of the model and ethical discussion

In the learning material the students have worked with data that were recorded with the same positioning of the smartphone. In practice a major challenge of human activity recognition is the position and orientation of the sensors on the user's body. To test how well the developed classification algorithm handles this problem, additional data were recorded. In one setting, the smartphone was in the user's hand, and in the other setting, it was stored in a backpack. The classification results for this data are significantly worse. Therefore, the need for sufficiently diverse training data is discussed. In addition, the possibility of incorporating additional sensors to tackle the bad results caused by different sensor positions is addressed. For example, data from a gyroscope and a magnetometer are available via phyphox and could be included.

The learning module is concluded with a discussion on different applications and use cases of human activity recognition. Critical aspects and possibilities of misuse are also discussed (e.g., data protection and privacy issues as well as the misuse of data for surveillance purposes).

First experiences with students

The learning module was piloted in the context of a mathematical modeling day (4–5 h) organized within the student program CAMMP⁴ (Schönbrodt et al., 2022b, p. 154). Nine students from grade 9 to 12 (corresponding to an age range of approx. 14–17) participated. The students had voluntarily registered for the project day, so it can be assumed that they were above-average motivated. Due to the small number of participants, we can only highlight first impressions. During the implementation, the observations were recorded. Following the modeling day, students completed a written questionnaire.

The very lively discussions and the final evaluation showed that the students had great interest in the topic of AI. In addition, all students confirmed that they had gained a better understanding of mathematical modeling and that they had developed a more profound understanding of the mathematical background of AI systems. One student expressed that she/he would have liked more independent programming (Hoeffler, 2022, pp. 80–88).

Summary and outlook

Due to their interdisciplinary nature, applications, and methods from the field of AI offer an excellent starting point for the design of interdisciplinary learning environments. This paper used the example of human activity recognition and one AI method to show how the three disciplines of mathematics, computer science and physics can be combined in an authentic way. Physics played a central role in understanding the problem, especially the sensor data. Mathematics was crucial in developing a mathematical model for classification and in using statistical measures to evaluate the classification results. Computer science came into play in the visualization and processing of the data and the implementation of the mathematical model.

Depending on the application under consideration, the integration of other subjects, both STEM and non-STEM, would be appropriate instead of physics. Knowledge from mathematics and computer science, however, plays a central role in (almost) all AI applications. Therefore, AI education fundamentally offers the chance to encourage mathematical as well as computer science learning processes. In addition to the kNN algorithm, the mathematical background of many other AI methods (including decision trees, the Support Vector Machine, factorization models, and neural networks) can be suitably didactically reduced and understood based on high-school mathematical knowledge, as underlined by Biehler & Fleischer (2021), Schönbrodt et al. (2022a), and Kindler et al. (in press).

In the current learning module, students work with acceleration data that has been recorded in advance. In the long run, the learning module should allow the flexible integration of the students' own data recorded with phyphox. Beyond that, the learning module is of a rather guided nature. It

⁴ CAMMP stands for Computational and Mathematical Modeling Program (see www.cammp.online/english/index.php)

would be interesting to implement the problem in the context of completely open mathematical modeling projects, for example in the context of modeling weeks (Schönbrodt et al., 2022b, p. 155).

References

- Blum, W. (1985). Anwendungsorientierter Mathematikunterricht in der didaktischen Diskussion? [Application-oriented mathematics teaching in the didactical discussion?] In Kahle et al. (Eds.), *Mathematische Semesterberichte. Zur Pflege des Zusammenhangs zwischen Schule und Universität*, 32(2), 195–232. Vandenhoeck & Ruprecht.
- Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, 43, 133–142.
- English, L.D. (2022). Interdisciplinary Mathematical Modeling. In Michelsen, C., Beckmann, A., Freiman, V., Jankvist, U.T., Savard, A. (Eds.), *Mathematics and Its Connections to the Arts and Sciences (MACAS)* (pp. 163–177). Springer. https://doi.org/10.1007/978-3-031-10518-0_9
- Hazzan, O., & Mike, K. (2022). Teaching core principles of machine learning with a simple machine learning algorithm: the case of the KNN algorithm in a high school introduction to data science course. *ACM Inroads*, 13(1), 18–25. <https://doi.org/10.1145/3514217>
- Hoeffler, K. (2022). Aktivitätserkennung auf dem Smartphone: Entwicklung von Unterrichtsmaterial für computergestützte mathematische Modellierungsprojekte [Activity recognition on smartphones: developing teaching materials for computer-based mathematical modeling projects], master thesis.
- Kindler, S., Schönbrodt, S., & Frank, M. (in press). Von der Schulmathematik zu künstlichen neuronalen Netzen [From school mathematics to artificial neural networks]. Beiträge zum Mathematikunterricht 2022.
- Mohsen, S., Elkaseer, A., & Scholz, S. G. (2022). Human Activity Recognition Using K-Nearest Neighbor Machine Learning Algorithm. In S. G. Scholz, R. J. Howlett & R. Setchi (Eds.), *Sustainable Design and Manufacturing* (pp. 304–313). Springer.
- Roehrig, G. H., Moore, T.J., Wang, H. H., & Park, M. S. (2012). Is adding the E enough? Investigating the Impact of K-12 engineering standards on the implementation of STEM Integration. *School Sciences and Mathematics*, 112(1), 31–44. <https://doi.org/10.1111/j.1949-8594.2011.00112.x>
- Schönbrodt, S., Camminady, T. & Frank, M. (2022a). Mathematische Grundlagen der Künstlichen Intelligenz im Schulunterricht [Mathematical foundations of artificial intelligence in the school classroom]. *Math Semesterberichte*, 69, 73–101. <https://doi.org/10.1007/s00591-021-00310-x>
- Schönbrodt, S., Wohak, K., & Frank, M. (2022b). Digital Tools to Enable Collaborative Mathematical Modeling Online. *Modelling in Science Education and Learning*, 15(1), 151–174, <https://doi.org/10.4995/msel.2022.16269>
- Staacks, S., Hütz, S., Heinke, H., & Stampfer, C. (2018). Advanced tools for smartphone-based experiments: phyphox. *Physics Education*, 53(4), 1–6. <https://doi.org/10.1088/1361-6552/aac05e>