Extended Abstract for the
Symposium on Integrating AI and Data Science into School Education Across Disciplines 2025

UNSUPERVISED MACHINE LEARNING AS LEARNING CONTENT IN LOWER SECONDARY SCHOOL

Katharina Bata[1] & Martin Frank[2]
[1]Karlsruher Institute of Technology, Germany, katharina.bata@kit.edu
[2]Karlsruher Institute of Technology, Germany

Focus Topics: Learning Materials, Explanatory Models

Introduction

The results of machine learning (ML) are becoming increasingly influential in everyday life, manifesting in applications such as recommendation systems that tailor content to individual preferences, or automated image recognition used in different contexts. This growing influence underscores the importance of integrating ML into school curricula, enabling students to critically engage with and evaluate related technologies (Tedre, 2021).

There are different approaches, suitable for different tasks and types of data, that are commonly referred to as ML-methods. One important approach is unsupervised learning, where data is analysed without pre-assigned labels or specific goals, as opposed to supervised learning, where the system is trained on labelled data (Bishop, 2016). The following paper discusses the usefulness of and theoretical foundations for teaching unsupervised learning, especially distance- and density-based clustering methods, in German lower secondary schools.

ML as learning content in K12

Research on ML teaching and learning has gained traction in recent years, especially for the target group of K12-students. Reviews show that students of different ages can grasp basic ML concepts, algorithms, and tasks (Martins & Gresse von Wangenheim, 2022; Marques et al., 2020) and that recent research has focused mainly on identifying pedagogical approaches, developing curricula, and creating tools to support ML education in K-12 settings (Sanusi et al., 2023). Given their conceptual overlap with computer science, statistics, and mathematics, ML methods are often introduced in mathematic or informatic or with reference to these subjects. Examples of this are AI4K12 (Touretzky et al., 2023), ProDaBi (Biehler & Schulte, 2018) and CAMMP (Schönbrodt et al., 2021).

Researchers who have studied ML as educational content often argue that ML should not necessarily be taught as a black box (Hazzan & Koby, 2022; Biehler & Schulte, 2018). A black box approach means that the use or quality of the relationships between the input and output of a ML-method is addressed without targeting how the method works between the input and output. The white box approach, on the other hand, aims to understand how the method works.

Distance- and density-based clustering

Clustering, a fundamental unsupervised ML technique, partitions a dataset into groups, or clusters, ensuring that points within the same cluster are more similar to each other than to those in different clusters. This similarity is typically quantified based on spatial proximity or density properties (Bishop, 2016). Distance-based clustering methods, such as k-Means, group data points by minimizing the distances between points within a cluster while maximizing the distances between clusters. In contrast, density-based methods, such as DBSCAN, define clusters as contiguous regions of high point density separated by low-density areas. Clusters are identified based on the local density of points, enabling the detection of arbitrarily shaped clusters and the handling of noise or outliers.

The description gives an indication of what can be confirmed by a more in-depth analysis of the content: The mathematical foundations of the selected methods (k-Means and DBSCAN) of unsupervised learning exclusively contain mathematical concepts that can already be addressed in secondary school math's lessons. Examples of this are distances or density characterizations. Discussing the methods with students is not only possible, but can thus be used as motivation for various mathematical concepts and at the same time promote ideas from ML as generating information from data.

In S. Podworny & S. Schönbrodt (Eds.), *Towards Fostering AI and Data Science Literacy in Schools Across Disciplines. Proceedings of the 1st Symposium on Integrating AI and Data Science into School Education Across Disciplines (AIDEA1 2025)*, Salzburg, Austria.

Unsupervised ML-methods have been less explored than supervised methods, which are more commonly integrated into teaching units (Marques et al., 2020). Therefore, the formulation of learning objectives and basic educational concepts is lacking for unsupervised learning methods as a learning object. With the aim of addressing the selected ML-methods in a workshop in school that focuses on the mathematical prerequisites of the pupils, these learning objectives and basic educational concepts should be approached in a design research project.

Theoretical foundation of a workshop on unsupervised machine learning

To design a workshop following a white box approach to unsupervised machine learning for the mentioned target group two aspects are crucial. The first one is a subject-related didactically oriented analysis of the learning content with regard to the underlying mathematical concept, the second one is the formulation of design principles (DP) for the learning activities in the workshop (Hußmann & Prediger, 2016).

The first step in the subject-specific didactic analysis is identifying the mathematical core of the method. This includes, for example, the different distance or density characterisations and the evaluation of the clusters using different similarity measures and optimisation criteria. The mathematical content is then structured with regard to the basic ideas and objectives in the area of unsupervised learning and reflected with regard to the prerequisites of the respective learning group. Based on the analysis, concrete learning objectives can then be defined, on the one hand with regard to mathematics and on the other hand with regard to unsupervised learning.

The DPs are chosen to suit the circumstances of the learning situation and to support the stated learning objectives. In relation to the project presented, DPs are formulated to activate students in school workshops, as well as considerations for the selection of appropriate examples and data sets for cluster analysis (Garfield & Ben-Zvi, 2009). Additionally, Jupyter Notebooks - a browser-based programming environment supporting Python (Barba et al., 2019) - are introduced as a tool for interactive, exploratory learning. The workshop will be based on digital worksheets created with Jupyter Notebooks enable exploratory work with data and the use of interactive elements or step-by-step support.

Discussion

The need to address machine learning concepts at school presents teachers and researchers with a major challenge (Sanusi et al., 2023). It turns out that the mathematical foundations of the methods offer a fruitful opportunity to find points of contact for students and, accordingly, mathematics lessons can be a way to address such concepts (Schönbrodt & Oldenburg, accepted).

The talk at AIDEA presents such an approach to unsupervised learning and gives insights into a workshop to make unsupervised machine learning accessible to German secondary school students. The theoretical considerations presented will form the foundation for the ongoing development of the workshop as part of a design research project. In the design research project, some learners are observed while working with the teaching learning material of the workshop in so called design experiments (Plomp & Nieveen, 2013). The data collected in the design experiments will be used to empirically investigate the developed teaching learning materials and thereby gain insights into the individual learning paths of the students, for example with regard to the following questions:

- *Which design principles are suitable to develop teaching and learning materials for secondary school students to support them in learning the concepts of unsupervised machine learning, in particular cluster analysis?*
- *Which hurdles do the secondary school students face when working with the developed teaching learning material on unsupervised machine learning, in particular cluster analysis?*

Addressing these research questions will guide the iterative refinement of the workshop content and materials, tailoring them to students' needs and exploring the accessibility of machine learning through its mathematical foundations.

References

Barba, L. A., Barker, L. J., Blank, D. S., Brown, J., Downey, A. B., George, T., Heagy, L. J., Mandli, K. T., Moore, J. K., Lippert, D., Niemeyer, K. E., Watkins, R. R., West, R. H., Wickes, E., Willing, C., & Zingale, M. (2019). *Teaching and Learning with Jupyter*. https://jupyter4edu.github.io/jupyter-edu-book/

Bishop, C. M. (2016). *Pattern recognition and machine learning*. Springer.

Biehler, R., & Schulte, C. (2018). Perspectives for an interdisciplinary data science curriculum at German secondary schools. In R. Biehler, L. Budde, D. Frischemeier, B. Heinemann, C. Schulte, S. Podwory & T. Wassong (Hrsg.), *Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts* (S. 1–14).

Garfield, J. B., & Ben–Zvi, D. (2009). Helping Students Develop Statistical Reasoning: Implementing a Statistical Reasoning Learning Environment. *Teaching Statistics*, *31*(3), 72–77. https://doi.org/10.1111/j.1467-9639.2009.00363.x

Hazzan, O., & Koby, M. (2022). Machine Learning for Non-Majors: A white box approach. *Statistics Education Research Journal*, *21*(2), 10.1–10.22. https://doi.org/10.52041/serj.v21i2.45

Hußmann, S., & Prediger, S. (2016). Specifying and Structuring Mathematical Topics. *Journal für Mathematik-Didaktik*, *37*(S1), 33–67. https://doi.org/10.1007/s13138-016-0102-8

Marques, L. S., Gresse von Wangenheim, C., & Hauck, J. C. R. (2020). *Teaching Machine Learning in School: A Systematic Mapping of the State of the Art. Informatics in Education*, *19*(2), 283–321. https://doi.org/10.15388/infedu.2020.14

Martins, R.M., & Gresse von Wangenheim, C. (2022). Findings on Teaching Machine Learning in High School: A Ten - Year Systematic Literature Review. *Informatics in Education*, *22*(3), 421-440. https://doi.org/10.15388/infedu.2023.18

Plomp, T., & Nieveen, N. (Hrsg.). (2013). *Educational Design Research: Part A: An Introduction*. Netherlands Institute for Curriculum Development.

Sanusi, I. T., Oyelere, S. S., Vartiainen, H., Suhonen, J., & Tukiainen, M. (2023). A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies*, *28*(5), 5967–5997. https://doi.org/10.1007/s10639- 022-11416-7

Schönbrodt, S., Camminady, T., & Frank, M. (2021). Mathematische Grundlagen der Künstlichen Intelligenz im Schulunterricht: Chancen für eine Bereicherung des Unterrichts in linearer Algebra [Mathematical foundations of artificial intelligence in the classroom: opportunities for enriching the teaching of linear algebra]. *Mathematische Semesterberichte*, *69*(1), 73–101.

Schönbrodt, S., & Oldenburg, R. (accepted). Mathematische Grundlagen als Schlüssel zu einem allgemeinbildenden Verständnis von KI: Theoretische Perspektiven und praktische Unterrichtsidee**n** [Mathematical foundations as a key to a general educational understanding of AI: Theoretical perspectives and practical teaching ideas]. *Mathematik im Unterricht*.

Tedre, M., Toivonen, T., Kahila, J., Vartiainen, H., Valtonen, T., Jormanainen, I., & Pears, A. (2021). Teaching Machine Learning in K–12 Classroom: Pedagogical and Technological Trajectories for Artificial Intelligence Education. *IEEE Access*, 9, 110558–110572.

Touretzky, D., Gardner-McCune, C., & Seehorn, D. (2023). Machine Learning and the Five Big Ideas in AI. *International Journal of Artificial Intelligence in Education*, *33*(2), 233–266. https://doi.org/10.1007/s40593-022-00314-1