"Art Heist at the Museum": Data Science Education through Educational Breakouts in an Informatics Teaching and Learning Lab

Kira Klaner[0009-0005-2851-1448], Nils Pancratz[0000-0001-7358-4148]

University of Hildesheim, Computer Science Education Research Group

Abstract. In an era of increasing digitalisation, data literacy is essential for responsible participation in society. However, key data science skills such as data collection, exploration and data-driven modeling and evaluation remain underrepresented in current curricula and teacher training. Educational Escape Rooms (EERs) offer a promising way to address this gap by embedding technical content into immersive, narrative scenarios. This paper presents an EER on Data Science, developed within the teacher education programme at the Informatics Teaching and Learning Lab at the University of Hildesheim. Framed as a fictional crime investigation, the EER uses the open-source data mining tool Orange to introduce supervised and unsupervised machine learning techniques, including decision tree classification and clustering. The scenario also fosters critical reflection on algorithmic bias through the use of AI-generated images and deliberately stereotyped character profiles.

Keywords: Data Science · Education · Educational Escape Room

1 Educational Objectives and Development Context of the Educational Breakout "Art heist at the museum"

Advancing digitalisation and the resulting "digitality" have made data ubiquitous across social networks, daily decisions, and economic processes. This trend drives a growing need for data literacy and a deeper understanding of data-driven technologies. *Data Science* (DS), understood as the practice of learning from data [1], is an interdisciplinary field closely linked to Computer Science (CS) through its reliance on Machine Learning (ML) approaches [3]. A basic understanding of DS is a prerequisite for responsible participation in a digital society [3].

Key DS competencies align with the data lifecycle [4, 9], including data collection (e.g., distinguishing primary from secondary sources), data exploration (e.g., pattern identification and data quality assessment), and data-driven modeling and evaluation (e.g., building predictive models using supervised learning). Despite their importance for informed participation in an increasingly data-driven society [3], these competencies are scarcely represented in current CS curricula and teacher education. Consequently, there is a pressing need for innovative teaching methods that foster both technical skills and learner motivation.

Educational Escape Rooms (EERs) – also referred to as *Educational Breakouts* – offer a promising approach by embedding data science learning objectives within immersive, time-limited scenarios requiring problem-solving and teamwork [11]. Designed with clear educational goals and specific target groups, EERs enhance cognitive and emotional engagement through narrative structures and collaborative challenges [11]. In recent years in particular, various EERs have been published in the STEM field [6]. For data science education in particular, EERs are especially well suited, as they offer a playful yet practice-oriented way to engage with key concepts such as data analysis, visualisation, and pattern recognition [8].

The EER presented in this poster contribution was developed as part of the CS teacher education programme at the University of Hildesheim "Informatics Teaching and Learning Lab" and tested with secondary school students and prospective teachers[1]. The EER focuses on supervised learning through decision tree learning (DTL). DTL has been successfully broken down into comprehensible learning materials on classification tasks such as distinguishing biting from non-biting apes [7]. Within the breakout narrative, decision trees help narrow down potential suspects.

2 Design of the Educational Breakout

A central element of the activity is the use of the open-source data analysis software "Orange", which offers a low-threshold and versatile introduction to data science concepts through its visual interface. "Orange" enables both novice and experienced users to construct data processing pipelines for the application of fundamental supervised learning techniques, including classification and regression [5], as well as for exploratory pattern recognition via cluster analysis within unsupervised learning frameworks [2].

In the EER, "Orange" is used to analyse realistic data sets embedded in a narrative context. Learners first used DTL to classify which individuals were likely suspects (e.g. 'criminal' vs. 'non-criminal') and narrowed down the group. They then compared the shopping data of the remaining suspects using clustering and identified the offender. The necessary data was provided by a fictional "detective" character, supported by additional materials such as newspaper articles and explanation videos. The learners not only operated the software, but also familiarised themselves with the most important algorithmic principles presented in the accompanying materials. These included concepts such as the distinction between training and test data and how classification and clustering algorithms work. The aim was to promote not only tool-specific skills, but also a fundamental understanding of data-driven decision-making processes.

After completing the EER, there was also a reflection phase in which ethical concerns and socio-cultural prejudices regarding AI-generated content were discussed. One example was an AI-generated reinterpretation of the Mona Lisa,

[1] Course material is published under https://lehr-lern-labor-inf.hosting.uni-hildesheim.de/materials/kunstraub-im-museum/

which shows a hyperfeminized figure with heavy make-up and a plunging neckline. This reflects the prevailing patterns of sexualization and objectification of women in both traditional media and AI-generated images [10]. The profiles of the suspects in the breakout scenario were intentionally designed with stereotypes – for example, the identified perpetrator was a tattooed person, reinforcing visual biases often associated with criminality. These elements were intentionally included to stimulate discussion about how algorithmic systems can not only reflect but also reinforce social stereotypes, particularly those related to gender, appearance or ethnicity. By integrating these examples into the EER, learners were encouraged to critically reflect on the ethical implications of algorithmically generated imagery, as well as on the responsibility of developers and users in addressing bias in algorithmic systems.

3 Implementation, Observations, Outlook, and Discussion

During the implementation of the EER, learners showed a high level of motivation, engagement and autonomy, especially in the 10-14 age group. The interactive storyline, hands-on materials and investigative structure enabled a strong emotional connection to the learning content and encouraged deep involvement and sustained attention throughout the activity. The format proved to be a highly effective introduction to the practical applications of DS. While scientific evaluation of the EER is yet to be conducted, experience showed learners' enthusiasm not only for solving the puzzles, but also for understanding the underlying data techniques – especially when these were directly linked to decision making in the game. However, some critical considerations also appeared during the implementation. The design of EER is not well suited as an introduction to DS. Central concepts such as clustering or DTL require prior knowledge or at least a simultaneous didactic preparation. In addition, the software Orange – although visual and intuitive – still requires a certain level of digital literacy and basic familiarity with data structures and attributes. In light of these observations, the EER is best suited for securing, applying or expanding previously introduced content. Embedded in a broader teaching sequence with corresponding preparation phases, this playful approach could therefore be more effective.

The presented EER demonstrates the potential of EERs as a powerful tool for combining technical skill acquisition with critical reflection on data-driven systems. By embedding supervised and unsupervised learning within a narrative context and integrating ethical discussions around AI-generated bias, the format not only fosters cognitive engagement but also supports socio-critical thinking, which is an essential component of data literacy.

However, in order to fully exploit the potential of the format, both the limitations and the possibilities for further development should be taken into account. For example, the EER is primarily suitable for applying, consolidating or deepening content that has already been introduced, but less as an independent introduction to the topic of DS. Future implementations should be more strongly embedded in didactically prepared teaching sequences that include both an introduction to the content and a structured follow-up. The developed EER also offers opportunities for interdisciplinary and inclusive educational approaches. Due to its cooperative, visual and interactive nature, the EER is particularly suitable for different target groups, such as secondary school pupils and teacher training students. The approach supports outreach projects by easing access to complex STEM content and reducing barriers in informatics, with particular benefit for programs promoting gender equality in CS education. Addressing bias in content and access contributes to a more inclusive and reflective digital education.

References

1. Cleveland, W.S.: Data science: An action plan for expanding the technical areas of the field of statistics. International Statistical Review **69**(1), 21–26 (2001). https://doi.org/10.2307/1403527
2. Demšar, J., Zupan, B.: Hands-on training about data clustering with orange data mining toolbox. PLOS Computational Biology **20**(12), 1–9 (2024). https://doi.org/10.1371/journal.pcbi.1012574
3. Engel, J.: Statistical literacy for active citizenship: A call for data science education. Statistics Education Research Journal **16**(1), 44–49 (2017). https://doi.org/10.52041/serj.v16i1.213
4. Grillenberger, A., Romeike, R.: Key concepts of data management: an empirical approach. In: Proceedings of the 17th Koli Calling International Conference on Computing Education Research. p. 30–39. ACM, New York (2017). https://doi.org/10.1145/3141880.3141886
5. Ishak, A., Siregar, K., Aspriyati, Ginting, R., Afif, M.: Orange software usage in data mining classification method on the dataset lenses. IOP Conference Series: Materials Science and Engineering **1003**(1), 012113 (2020). https://doi.org/10.1088/1757-899X/1003/1/012113
6. Lathwesen, C., Belova, N.: Escape rooms in stem teaching and learning—prospective field or declining trend? a literature review. Education Sciences **11**(6), 308 (2021). https://doi.org/10.3390/educsci11060308
7. Michaeli, T., Seegerer, S., Kerber, L., Romeike, R.: Data, trees, and forests - decision tree learning in k-12 education. In: Kinnaird, K.M., Steinbach, P., Guhr, O. (eds.) Proceedings of the Third Teaching Machine Learning and Artificial Intelligence Workshop. Proceedings of Machine Learning Research, vol. 207, pp. 37–41. PMLR (19–23 Sep 2023)
8. Nelson, V., Crea, J.: The data science instructional escape room—a successful experiment. CHANCE **34**(2), 53–58 (2021). https://doi.org/10.1080/09332480.2021.1915034
9. Olari, V., Romeike, R.: Data-related concepts for artificial intelligence education in k-12. Computers and Education Open **7** (2024). https://doi.org/10.1016/j.caeo.2024.100196
10. Sandoval-Martín, T., Martínez-Sanzó, E.: Perpetuation of gender bias in visual representation of professions in the generative ai tools dall·e and bing image creator. Social Sciences **13**(5), 250 (2024). https://doi.org/10.3390/socsci13050250
11. Veldkamp, A., van de Grint, L., Knippels, M.C.P., van Joolingen, W.R.: Escape education: A systematic review on escape rooms in education. Educational Research Review **31**, 100364 (2020). https://doi.org/10.1016/j.edurev.2020.100364