

Invitation

Paderborn Colloquium on Artificial Intelligence and Data Science Education at School Level

Organized by the ProDaBi project (<https://www.prodabi.de/en/>), Paderborn University, Germany

Project directors: Rolf Biehler (statistics and mathematics education) & Carsten Schulte (computer science education)

Team: Yannik Fleischer, Daniel Frischemeier, Lukas Höper, Sven Hüsing, Susanne Podworny

Data science, artificial intelligence, machine learning, data literacy, and statistical literacy concerning secondary education are currently discussed in the communities of scientists and educators in statistics, mathematics, computer science, social and natural sciences, and media education. Our colloquium intends to bring together these perspectives and communities to create an interdisciplinary community for scientific exchange.

Since data science and artificial intelligence have become more and more relevant in industrial and economical automation processes, marketing processes, and monitoring in politics, both topics permeate nearly all areas of life. These influences raise questions about future possibilities for social participation, self-determination, and self-realization in the professional and private sector, resulting in the need for educational processes that address these issues in school. For the teaching of mathematics and computer science completely new challenges have emerged, as well as for the subjects of the socio-scientific field and cross-curricular media education.

In our colloquium, we want to take up these issues and discuss the state-of-the-art and future trends of artificial intelligence and data science education at school level. We also want to discuss fundamental ideas of data science as they are conceptualized by experts in this field since a broad perspective of data science as a scientific discipline is needed to inform curriculum development. Contributions to the colloquium will also present practice-oriented research as well as research on teachers' professional development.

ProDaBi (Project Data Science and Big Data at School) develops research-based teaching material and professional development courses for teaching data science and artificial intelligence for grades 5 to 12. It was initiated and is funded by the Deutsche Telekom Stiftung since 2018.

Program

27th October 2021, 4 p.m. to 6:30 p.m. (CEST, UTC+2)

- *Data detective clubs in the time of COVID-19*
Jan Mokros and **Bill Finzer** (USA)
- *Teaching machine learning in school: Some emerging research trajectories*
Matti Tedre and **Henriikka Vartiainen** (Finland)

24th November 2021, 4 p.m. to 6:30 p.m. (CET, UTC+1)

- *Beyond bias. Locating questions of injustice in data science and artificial intelligence*
Tobias Matzner (Germany)
- *Bringing together statistics and computer science education: Machine learning by decision trees grounded in students' data exploration experiences*
Rolf Biehler and **Yannik Fleischer** (Germany)

12th January 2022, 4 p.m. to 6:30 p.m. (CET, UTC+1)

- *Learning data science through civic engagement with open data*
Graham Dove (USA)
- *Why should students take a data science course?*
Rob Gould (USA)

Each presentation will be followed by an intensive discussion aiming at community building.

The colloquium is open and free for everyone and will be held via zoom. Please register at <https://www.prodabi.de/registration-colloquium>.

Please distribute the information to interested colleagues so that they can also register.

Information about presentations and speakers

Data detective clubs in the time of COVID-19

Jan Mokros and Bill Finzer (USA)

Session 1, part 1: October 27, 2021, 4:00 p.m. (CEST, UTC + 2)

The COVID-19 pandemic presents an opportunity to engage young people in exploring how data can be used to understand a public health crisis, make decisions, and save lives. In this session we will describe a multifaceted project involving an adventure story about COVID-19 that is connected to data challenges in which CODAP (Common Online Data Analysis Platform) is used to explore time series of pandemic data. This work takes place in out-of-school clubs around the US, comprising 20 hours over two to three months with students who are 10-14 years old.

We will focus on two aspects of this work: First, we'll demonstrate and discuss the affordances of CODAP and accompanying datasets in understanding how a dynamic pandemic unfolds. For example, CODAP interacts well with NetLogo, which means students (and those attending our session) can set parameters for infectivity and run multiple simulations to see how many people get sick, how many recover, and how long the outbreak lasts. In addition, CODAP's "Story Builder" feature enables youth to combine graphs, photos, and text to tell the story of what happens over time with COVID under different circumstances and in different places.

Second, we'll discuss the challenges and opportunities of working with real-time, highly relevant data that transcend the boundaries of school curricula. Most students do not study epidemiology in secondary school, though the subject area is an ideal vehicle for learning about data. Students' work with data from pandemics also integrates social science, public policy, and the science of viruses.

The session will conclude with a discussion of the social-emotional aspects of using sensitive data. COVID data, like most data that truly matter, elicit a range of social and emotional issues, and we believe it is part of our role as data science educators to address these concerns.

Speakers:



Dr. Jan Mokros is a developmental psychologist who is currently directing National Science Foundation-funded projects focused on how youth learn about data in out-of-school clubs. Jan's work with data science education introduces youth to topics including Lyme disease, teens' use of time, sports injuries, and COVID-19. The COVID project centers on using a combination of CODAP, data activities, and a young adult adventure book, "The Case of the COVID Crisis", by Pendred Noyce, to explore infectious disease epidemics. Afterschool programs around the US are using this program with youth who are underserved with respect to STEM. Jan is a Senior Research Scientist at Science Education Solutions.

In prior positions, she has designed curriculum and conducted research at TERC and at the Maine Mathematics and Science Alliance. She has authored three books, including one for museum educators on incorporating math into exhibits and programs, and one for parents on exploring math in everyday life. She has been involved as a writer and researcher for the math curriculum Investigations in Number, Data and Space.

Bill Finzer's work has long centered on getting students using data in every subject they study. He led the Fathom Dynamic Data Software development team at KCP Technologies before joining the Concord Consortium in 2014 where he leads development of the Common Online Data Analysis Platform (CODAP). He has been a classroom teacher, curriculum developer, teacher professional development course designer and leader, and educational software developer. Bill works with staff of many projects both inside and outside Concord Consortium to help them make use of CODAP. He loves nothing better than fixing bugs and implementing new features.



Teaching machine learning in school: Some emerging research trajectories

Matti Tedre and Henriikka Vartiainen (Finland)

Session 1, part 2: October 27, 2021, 5:10 p.m. (CEST, UTC + 2)

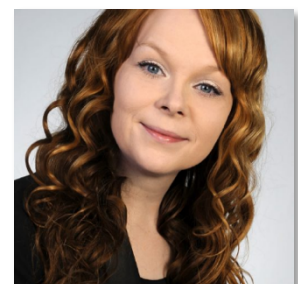
A major technological shift has recently triggered discussions about the need to amend computing education at all education levels. Traditional, rule-based automation has been joined by machine learning (ML), which, when provided with enough computing power and data, has enabled new classes of jobs to be automated, and thus expedited automation in the society, workplace, and in people's everyday lives. Although ML has become an integral part of our lives, communities, and societies, it has gained very little attention in K–12 (school) computing education which mainly focuses on rule-based programming and computational thinking. This talk will map the emerging trajectories in educational practice, theory, and technology related to teaching machine learning in K-12 education. It will situate that research in the broader context of computing education and describe what changes ML necessitates in the classroom. The talk will outline the paradigm shift that will be required in order to successfully integrate machine learning into the broader K-12 computing curricula.

Speakers:



Dr. Matti Tedre is a professor of computer science, especially computing education and the philosophy of computer science, at the University of Eastern Finland. His 2019 book "Computational Thinking" (The MIT Press, with P.J. Denning) presented a rich picture of computing's disciplinary ways of thinking and practicing, and his 2014 book "Science of Computing" (Taylor & Francis / CRC Press) portrayed the conceptual and technical history of computing as a discipline.

Dr. Henriikka Vartiainen is a senior researcher at the University of Eastern Finland, School of Applied Educational Science and Teacher Education. Currently, her research focuses especially on learning Machine Learning through co-design as well as on the ways to support children's data agency.



Beyond bias. Locating questions of injustice in Data Science and Artificial Intelligence

Tobias Matzner (Germany)

Session 2, part 1: November 24, 2021, 4:00 p.m. (CET, UTC + 1)

Data Science and Artificial Intelligence have repeatedly come under scrutiny because of injustices they produce. AI based services only work for a certain part of the populace, scoring and ranking systems are skewed towards certain groups, errors concentrate on specific persons etc. Thus, when teaching Data Science and Artificial Intelligence, such issues should be part of the curriculum. Most often in research and in educational concepts, injustice is discussed as bias outputs of information processing. In social studies of technology, recently some people have begun to challenge this idea of equating injustice with bias. The talk will shortly present why this is the case – and how this challenge from current research opens up new ways of discussing the social impact of Data Science and Artificial Intelligence in schools.

Speaker:



Dr. Tobias Matzner is professor for “Media, Algorithms, and Society” at Paderborn University in Germany. His work combines theories of (digital) media and technologies with approaches from political philosophy, cultural studies, and social theory. He has studied philosophy and computer science in Karlsruhe, Rome, and Berlin. Prior to his appointment in Paderborn, he has been working at the International Centre for Ethics in the Sciences and Humanities at the University of Tübingen as well as the New School for Social Research in New York.

Bringing together statistics and computer science education: Machine learning by decision trees grounded in students’ data exploration experiences

Rolf Biehler and Yannik Fleischer (Germany)

Session 2, part 2: November 24, 2021, 5:10 p.m. (CET, UTC + 1)

Trees can be used to visualize decision rules for classifications, and students may have encountered trees for different purposes in the mathematics or computer science classroom already. Simple decision trees can support everyday decisions. A new idea for students is to use trees for predictive modeling (classification) in multivariate data sets. The human construction of trees has to be based on insights into the data and its context. Based on such experiences, algorithms for the automatic creation of trees can be developed and critically evaluated. Essential elements of predictive modeling such as the distinction between training and test data, overfitting, consequences of bias in the data (random or systematic sources), different evaluation criteria based on the confusion matrix are essential aspects to be taught.

We have developed material and educational guidelines for use at several educational levels (grade 5/6, 9/10, and 11/12). These materials include an unplugged decision game with data cards for use at all levels, but particularly in grades 5 and 6. We use different computational tools: CODAP

(codap.concord.org) as a web-based, easy-to-use data exploration tool with a plug-in for “manual” creation of decision trees based on data. For enabling automation, we designed different types of Jupyter Notebooks based on Python. The different types require different levels of coding skills from the students. The basic notebooks can be used for a larger group of students, but advanced notebooks are available for in-depth activities and higher educational levels. Basic Jupyter Notebooks appear menu-driven. In more advanced notebooks, students get worked examples for computational essays containing with explicit code, which they can adapt and enhance for their own data and predictive modeling problems. All notebooks use libraries for data exploration and for automated decision tree creation. We have adapted them for educational purposes from professional sources.

Students encounter various multivariate data sets. These include data on nutrition values of food, data on (social) media use of adolescents, and data from medicine on heart diseases. In addition, parking lot occupancy data from their town were used, where predictive modeling is applied to help reducing parking search traffic and related emissions.

We will present some of our materials for teaching decision trees and the first results from studies in the classroom where we used the material.

Speakers:



Dr. Rolf Biehler is professor for didactics of mathematics at Paderborn University. His research interests include probability, statistics and data science education, university mathematics education and the professional development of mathematics teachers. He was a co-founder and co-director of the Centre for Research in University Mathematics Education (www.khdm.de). He is engaged in the International Association of Statistics Education (IASE) and has worked as an editor or editorial board member for several international journals and books on mathematics education. He is currently co-directing the Project Data Science and Big Data at School (www.prodabi.de/en).

Yannik Fleischer is a PhD student in mathematics education research at Paderborn University, Germany. His main research interest is developing a conception for teaching machine learning methods in school with a focus on decision trees, and to evaluate this by developing and examining teaching materials in practice. Since 2019, he has been teaching year-long project courses on data science in upper secondary and developing, implementing, and evaluating teaching modules for different levels in secondary school, mainly about machine learning with decision trees.



Learning data science through civic engagement with open data

Graham Dove (USA)

Session 3, part 1: January 12, 2022, 4:00 p.m. (CET, UTC + 1)

In this talk I will discuss work undertaken for the project “Learning Data Science Through Civic Engagement With Open Data”. This project, which is funded through the National Science Foundation’s Advancing Informal STEM Learning (AISL) program, studies the informal data science learning that takes place within workshops, and other events and activities, that have been developed to support community engagement with civic open data in New York City (NYC). NYC is a leader in Open Data initiatives, which are centered around the NYC Open Data portal, and which have become enshrined in the City Charter. It also has a large and highly diverse population, including many traditionally underserved communities. As government service provision becomes increasingly digital, large amounts of data are generated and subsequently used to assess need, drive service delivery decisions, and evaluate effectiveness. Services producing these data include education, transport, and 311 service requests (a non-emergency municipal service available in many cities for reporting problems such as noise or public safety concerns), and the data they produce can be probed to ask many important questions such as: “How do City agencies respond to noise in my neighborhood?”, “How do waste and recycling services in my neighborhood compare with others?”, and “Are there more construction permits issued for my neighborhood than similar areas?”. To better understand how diverse communities might access and analyze these data to answer questions, share narratives about issues of concern, and respond to data driven policy and resource allocation, we are studying programs offered by the Mayor’s Office of Data Analytics (MODA) and BetaNYC. MODA is the NYC agency with overall responsibility for the City’s Open Data programs, while BetaNYC is a leading nonprofit organization working to improve the lives of NYC residents through civic design, technology, and engagement with government open data. We study the ecosystem that has emerged around the programs these organizations offer as a possible model for identifying, validating, and evaluating best practices; including questions of participation and potential barriers to entry.

Speaker:



Dr. Graham Dove is a human-computer interaction researcher, with experience in participatory approaches to design and citizen science. Based in NYU Tandon’s Dept. of Technology Management and Innovation (TMI), and the Center for Urban Science and Progress (CUSP), Graham investigates ways that people who are not experts in data science can use quantitative data and artificial intelligence to inform decision making, advocacy, and creativity in design. Current projects include investigating the informal learning that takes place around NYC Open Data, designing data rich interfaces to support future healthcare work practices. SONYC

(Sounds of New York City), which investigates approaches to monitoring and mitigating noise pollution. He has previously worked in Denmark and the UK.

Why should students take a data science course?

Rob Gould (USA)

Session 3, part 2: January 12, 2022, 5:10 p.m. (CET, UTC + 1)

The Mobilize Intro to Data Science (IDS, www.introdatascience.org) course was first offered in 2017 and was, at that time, the only data science course designed for secondary students in the U.S., and possibly anywhere. IDS was developed through a partnership between the Los Angeles Unified School District (the second-largest district in the U.S.), the UCLA Department of Computer Science, the UCLA Graduate School of Education and Information Sciences, and the UCLA Department of Statistics. The IDS curriculum has several special features. First, it relies on a data collection paradigm, Participatory Sensing (Burke, et al 2006), in which students use mobile devices to collect multivariate data together. Second, the curriculum teaches students to use R, via the interface Rstudio and the mobilizR package, to organize, prepare, and analyze data. Finally, it relies on a student-centered, activity-based pedagogy. The primary goal of IDS is to develop in students the ability to synthesize statistical and computational thinking (DeVeaux, et al. 2017; Gould 2021) in order to work, and live, ethically and productively in a data-driven world.

Since 2017, interest in preK-12 data science and “data literacy” has blossomed in the U.S (and elsewhere), and varying visions and purposes for data science education have emerged. Data science education is many things to many people; it can be seen as an approach to developing “data literacy”, improving programming skills, increasing “college readiness”, improving equity in mathematics (Burdman,2015), and developing students’ appreciation for and skill in for mathematics. While we agree that many of these purposes are welcome consequences of a well-designed data science course, in our vision of data science education these are secondary to the primary goal, which is to teach students to analyze complex, multivariate data and to develop what has been called “data acumen”. Many in the U.S. see secondary-level Data Science as a sub-discipline of mathematics (Levitt 2019), which naturally affects the scope and intent of a course.

In this presentation, we’ll describe the IDS vision of data science and IDS’s role in developing data literacy for secondary students. We’ll also discuss some of the challenges that remain in implementing this vision, including teacher preparation and university acceptance.

Speaker:



Dr. Rob Gould is a teaching professor and vice-chair of undergraduate studies in the Department of Statistics at UCLA, active in statistics and data science education since 1994. He is the founder of DataFest, a 48-hour undergraduate data analysis competition sponsored by the American Statistical Association and held at over 40 sites around the world. He is co-author with Colleen Ryan and Rebecca Wong of an introductory statistics book. He was the lead Principal Investigator of the NSF-funded Mobilize project, which produced the Introduction to Data Science curriculum, the first high school data science curriculum in the

U.S. Rob was elected Fellow of the American Statistical Association in 2012; in 2019 was awarded the CAUSE Lifetime Achievement Award for Statistics Education and the ASA Waller Distinguished Teaching Career Award. He received his B.S. from Harvey Mudd College in 1987 and PhD in Mathematics (concentration on Statistics) from the University of California, San Diego, in 1994. He is Vice-president of International Association of Statistics Education (IASE).