DATA SCIENCE EDUCATION K-12

Individual Showcase: What do high school students experience and learn during a two-day datathon?

Kathryn Jessen Eller, East Bay Educational Collaborative, kathryn.eller@ebecri.mailto:kathryn.eller@ebecri.org
Jillian Schneider, Bristol Warren Regional School District, jillian.schneider@bwrsd.org
Hamish Fraser, Brown Center for Biomedical Informatics, hamish_fraser@brown.edu
Jeremy Warner, Brown Lifespan Center for Clinical Cancer Informatics and Data, jeremy_warner@brown.edu
Ece Uzun Brown Lifespan Center for Clinical Cancer Informatics and Data, dilber_gamsiz@brown.edu
Sandeep Jain, Brown Lifespan Center for Clinical Cancer Informatics and Data <sandeep_jain@brown.edu>
Joao Carlos R G De Matos, Duke University Health System, joao.matos@duke.edu
Doug Rademacher, Metropolitan Regional Career and Technical Center, drademacher@metmail.org
Anushree Bopardikar, TERC, anushree_bopardikar@terc.edu
Michael Cassidy, TERC, michael_cassidy@terc.edu
Brandy B. Jackson, Scoutlier by Aecern, brandy@aeceern.com
Elizabeth Price, Scoutlier by Aecern, bethprice@aecern.com
Luk Henrik, Scoutlier by Aecern, luk@aecern.com
Judy Wawira Gichoya, Emory University School of Medicine, judywawira@emory.edu
Leo Anthony G Celi, Massachusetts Institute of Technology, lceli@mit.edu

Session Description

What do high school students learn from a two-day datathon during which they tackle data to visualize the impact of biased data on healthcare decisions? How do they interact with their team of high school students, data scientists, clinicians, and teachers? What did we, the developers and leaders of the datathon, learn? How would we approach it differently next year? Our goal is to answer these questions plus share lessons learned. We will then divide the audience into teams to brainstorm ways to approach and solve some of the problems we experienced and hopefully recruit some audience members to participate in our June 2025 Brown University Health Artificial Intelligence (AI) Systems Thinking for Equity (HASTE) Datathon in Providence, Rhode Island (Brown University Datathon, 2024).

Introduction

We explored how diverse high school students experienced a two-day datathon. The intent was to expose students to an authentic cross-disciplinary collaborative workforce experience within which they synthesized and applied what they had learned from a semester-long course, Data Science, AI, and You (DSAIY) in healthcare. The June 2024 datathon, highlights flaws in the "Human Dimensions of Data" by emphasizing the importance of critically assessing where, how, when, and from whom data is collected. We encouraged viewing strength in diversity and celebrated everyone's perspective by convening people across levels of education and disciplines. Our theoretical framework of context-based (Rose, 2012; Yu et al., 2015) experiential learning (Means & Stephens, 2021) is relevant across the preK-12 community because educators could apply the datathon model towards deepening student comprehension, synthesis, and application during integrated instruction and across other disciplines.

Project Context

Our datathon occurred in an informal educational setting at Brown University wherein students teamed with mentors to solve and discuss problems with bias in data. This datathon and associated course are part of the DSAIY program funded to explore a novel learning ecosystem for teaching data science and machine learning skills to high school students (NSF #2148451). Our aim is to understand what students experienced during a two-day datathon and how this impacted their thinking. Our primary challenge was engaging high school students within the datathon's complex, abstract learning environment.

Methods

A datathon planning committee assigned 46 high school students and 100+ mentors into 7-8 diverse groups. On Day 1, teams explored the impact of bias on mortality data from the 2020 Women in Data Science (WIDS) Conference (Lee et al., 2020). On Day 2, Dr. Leo Celi of Massachusetts Institute of Technology (MIT) led a HASTE workshop (Brown University Datathon, 2024) wherein participants learned about and discussed the impact and political

ramifications of Large Language Models. Examples included predicting race in medical imaging (Gichoya et al., 2022) and sexual orientation from faces (Wang and Kosinski, 2018).

The datathon occurred within a large open space equipped with Wi-Fi access and 10 round tables with a charging station and large monitor. We included speakers and "data leaders" trained by Dr. Celi and instructions within a GitHub Jupypter notebook. We report on data from a post-datathon feedback survey, interviewing student groups (coded) and individual participants, video documentation, and team presentations.

Transcripts of student datathon interviews were coded inductively and deductively. Main codes included general awareness about data science/AI/ML (including their intersection with healthcare/medicine); interest in related careers; specific skills/big ideas learned about; in-class preparation to engage with authentic data and professionals; and networking at the datathon.

Results

Brown University Datathon HASTE included 100+ data scientists, clinicians, and student (undergraduate, graduate, and medical school) mentors, plus 11 teachers from 3 Title I urban schools and one suburban school. Twenty participants responded to the post-datathon survey. Almost all (95%) reported they learned something new or valuable from the datathon experience, with a medical doctor or nurse commenting, "*It was great to engage in discussions on data science and network with the other professionals here*." Most would participate in this datathon again, reported they connected with people at the datathon, and that the event changed their thinking.

Testimonials:

High School Student: "*I learned many valuable things from the datathon. From AI applications to how machine learning came to be and even examples of both. It was an amazing experience to be a part of. It gave me an opportunity to explore deeper within the medical field. It also gave me the opportunity to talk to very intelligent people who are experts in their field. As a Freshman, that's pretty awesome. Overall, I loved the setup, especially the food, and the many opportunities this event gave me to learn more about AI and machine learning*".

Young Mentor Research Data Analyst: "*I would like to be invited to future (local) datathons. This was a great experience for me to learn more about predictive analytics and AI.*"

Teacher: "*My students engaged in profound discussions about the data, presentations, and the ethics of Artificial Intelligence. Those who attended only on the second day were pleased to integrate smoothly into the group and felt they could meaningfully contribute to the dialogue. Many were surprised by how much they learned. They were thrilled to be part of such a diverse group of experts, and their enthusiasm was palpable, with lively conversations about the datathon continuing during the bus rides to and from the event.*"

High school student interviews revealed that the datathon allowed them to learn more deeply about Python coding, computer science, data science, and technology. Although some students did not think the datathon fit with their career interests, they thought it was "*great to know things* [they] *will probably need in life*" and that networking with mentors in their team at the datathon was "*opening* [them up] *to a new experience*." Students recognized that the curriculum had prepared them for the datathon "*enough to not be completely lost*." For example, they could contextualize bias, data, and the basics of coding because these had been discussed previously in class.

Conclusion

We learned that a two-day informal event, such as a datathon, can impact student career interest and thinking. This implied that even short informal events that build upon classroom learning can affect the student experience. We will share our lessons learned to inform others interested in hosting their own high-impact datathon. We look forward to audience discussion and feedback to improve our next Datathon and invite them to join us in Providence, Rhode Island, for our next datathon in spring 2025.

References

Brown University Datathon | The Warren Alpert Medical School of Brown University https://medical.brown.edu/HASTE_Datathon Accessed: 2024-08-08

Gichoya, J. W., Banerjee, I., Bhimireddy, A. R., Burns, J. L., Celi, L. A., Chen, L. C., ... & Zhang, H. (2022). AI recognition of patient race in medical imaging: a modeling study. The Lancet Digital Health, 4(6), e406-e414.

Lee, M., Raffa, J., Ghassemi, M., Pollard, T., Kalanidhi, S., Badawi, O., ... & Celi, L. A. (2020). WiDS (women in data science) datathon 2020: ICU mortality prediction (version 1.0. 0). PhysioNet.

Means, B. M., & Stephens, A. (Eds.). (2021). Cultivating interest and competencies in computing: Authentic experiences and design factors. National Academies Press. https://doi.org/10.17226/25912

Rose, D. E. (2012). Context-based learning. In N. Seel (Ed.). Encyclopedia of the Sciences of Learning (pp. 799-802). New York: Springer.

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Journal of Personality and Social Psychology, 114(2), 246–257. https://doi.org/10.1037/pspa0000098

Yu, K. C., Fan, S. C., & Lin, K. Y. (2015). Enhancing students' problem-solving skills through context-based learning. International Journal of Science and Mathematics Education, 13(6), 1377-1401.