Abstract

Web-surfing histories, online newspapers, streaming music, and stock prices all show that we live in an age of data. Extracting meaning from data is necessary in many fields to comprehend the information flow. This need has fueled rapid growth in data science education aiming to serve the next generation of policy makers, data science researchers, and global citizens. Initially, teaching practices have been drawn from data science’s parent disciplines (e.g., computer science and mathematics). This work begins investigating data science education in its own right by aiming to identify preconceptions students may have when they first enter a data science classroom, and what other courses from related programs are shaping their preconceptions. The investigation will focus on topics identified in the National Academies of Sciences, Engineering, and Medicine (NASEM) Report: Data Science for Undergraduates: Opportunities and Options.

Main Objectives of Project

1. Identify student misconceptions, difficulties, and non-expert thinking about data science concepts.
2. Document data science concepts that engineering (and other) students develop outside of data science courses.
3. Develop a survey to discover gaps between what academic data science programs offer students and what the above data science concepts
4. Develop a protocol for conducting interviews with students to probe student misconceptions/knowledge
5. Discover what data science skills students have acquired from other courses.
6. Identify disconnects between core elements of data science curricula and what early career data science practitioners use in their daily work.

Timeline of Activities

<table>
<thead>
<tr>
<th>Year 1 2019</th>
<th>Spring</th>
<th>Spring</th>
<th>Summer</th>
<th>Fall</th>
<th>Year 2 2020</th>
<th>Spring</th>
<th>Spring</th>
<th>Summer</th>
<th>Fall</th>
<th>Year 3 2021</th>
<th>Spring</th>
<th>Spring</th>
<th>Summer</th>
<th>Fall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prerequisites/Industriy survey</td>
<td>Finalize collection &amp; difficulty protocol</td>
<td>Training: data science program mapping &amp; course performance</td>
<td>TAs and Instructor data collection training</td>
<td>Distributed, in vivo data collection</td>
<td>Distributed, in vivo data collection</td>
<td>Development of open-ended questions</td>
<td>Collected student thinking on open-ended questions</td>
<td>Development of open-ended questions</td>
<td>Collected student thinking on open-ended questions</td>
<td>Final reports and dissemination produced</td>
<td>Open-ended questions</td>
<td>Open-ended questions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Evaluation Plan

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Outcomes</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conduct Philosophy</td>
<td>Improve communication &amp; collaboration</td>
<td>Qualitatively score with PA and annual meeting with PA</td>
</tr>
<tr>
<td>Create Logic Model</td>
<td>Ensure that key project indicators and outcomes are being assessed</td>
<td>Conduct annual mapping from evaluation result to Logic Model indicators for NSF Report</td>
</tr>
<tr>
<td>Develop &amp; Administer Interview Protocol</td>
<td>Evaluate PI &amp; stakeholder engagement, attitudes and feedback on progress and approaches</td>
<td>Develop interview protocol for PIs and stakeholders</td>
</tr>
<tr>
<td>Develop &amp; Administer Examination Tools</td>
<td>Assess effectiveness of examination targets in engaging educators, increase understanding, and catalyze change in data science education</td>
<td>Develop professional development evaluation tool</td>
</tr>
<tr>
<td>Evalue Annual PI Meetings</td>
<td>Assess engagement of stakeholders</td>
<td>Report outcomes &amp; impact in final summative report(s)</td>
</tr>
<tr>
<td>Provide Annual Evaluation Report</td>
<td>Provide annual formative assessment and insights</td>
<td>Formative feedback assessment in first two years and provide annual reports</td>
</tr>
<tr>
<td>Provide Course corrections &amp; recommendations</td>
<td>Conduct summative assessment in year 1 and provide summative report</td>
<td></td>
</tr>
</tbody>
</table>

Broader Impacts

NASEM’s report Data Science for Undergraduates: Opportunities and Options recommends that “…institutions should … work together to develop professional approaches for evaluation, shared investment and evaluation frameworks, data sets, and culture … [and] align educational evaluation with market impacts” [Rev. 5.3] [2].

Summary of Broader Impacts

• Connecting academic lines of data science to industry lines of data science through an explicit collection of disconnects between the two
• Gathering common misconceptions will help focus the teaching of various data science concepts by illuminating which concepts are currently the most misunderstood by students
• Better understanding of common misconceptions will facilitate diversity, inclusion, and accessibility in the discipline
• Providing a necessary step towards creating a standardized assessment tools in data science could quantify the value of various data science educational programs
• A clearly defined scope for data science will support the continued creation of educational undergraduate opportunities (courses, programs, etc.) with potential impacts in graduate education
• Creating a platform that will allow for interdisciplinary collaborations and consensus for data science research (especially educational research and engineering education research)

Intellectual Merit

An extensive and repeated study of student misconceptions within data science education and its connection to data science education research. As a core topic in the scholarship of teaching and learning, this knowledge will help educators develop more effective instructional materials to commonly misconceptions. This follows the National Research Council’s (NRC) recommendation that “more studies measure outcomes other than test scores and course performance” [1].

Summary of Intellectual Merits

• Formal documentation and study of student misconceptions within data science
• Investigation of inter- and intra-disciplinary transfer of data skills
• Design and testing of a distributed, in vivo qualitative data collection method

References

2. (2012) Investigating the Data Science Degree: The Undergraduate Perspective. University of Texas at Dallas, School of Biomedical Engineering, Biomedical Informatics, and Data Science.