Pushing Data Science Education into the Real World

Daniel Turek, Anthony Suen, Dav Clark

ABSTRACT

The discipline of data science has been viewed as an convergence of high-power computing, data visualization and analysis, and data-driven application domains over the past decade. Prominent research institutions and private sector industry have been quick to embrace the importance of data science, but the foundations for effective tertiary-level data science education are conspicuously absent. This is nothing new, however, as the university has a well-established tradition of developing its educational mission hand in hand with the development of novel methods for human understanding (Feingold, 1991). Thus it is natural that universities "figure out" data science hand in hand with the development of needed pedagogy. We consider the development of data science education with respect to recent trends in interdisciplinary and experiential education, along with agile and design thinking methodologies to understand how they could apply to data science educational programs. This historical perspective motivates us to consider what factors are necessary to drive effective data science education, which range from a complete end-to-end workflow, technological tools for development and team communications, and appropriate motivation and incentives. The first iteration of the *Berkeley Institute for Data Science (BIDS)* Collaborative started in the University of California, Berkeley in the Spring of 2015 is used as a case study. From this we draw lessons learned and form a hypothesis regarding the necessary ingredients for effective data science education at the tertiary level. This hypothesis will be tested and revised in subsequent iterations of the BIDS Collaborative as we continue our study of effective data science education, research, and social impact.