Can colleges launch data science programs fast enough?

Higher ed has struggled to keep up with the surging demand for data scientists, but a new open-source curriculum from IBM could help.

By Natalie Schwartz

Published Nov. 26, 2019

When the University of California, Berkeley rolled out a new undergraduate data science program last fall, nearly 800 students immediately signed up for major, making it one of the most popular degrees on campus.

That move has positioned UC Berkeley to be a leader in producing data scientists, an in-demand and high-paying job that requires workers to be adept in topics such as statistics, machine learning and programming.

But colleges are largely struggling to add or expand their programs to keep up with the growing demand for such workers. More than 150,000 jobs requiring data science skills are unfilled, with particularly severe shortages in large cities such as New York City and Los Angeles, according to a 2018 report from LinkedIn.

"Data science is an incredibly hot area," said Susan Davidson, a computer and information science professor at the University of Pennsylvania, in an interview with Education Dive. "Yet there are relatively few places and resources that currently are available for teaching data science."

IBM is hoping to help. In September, the technology company announced it is working with the U of Pennsylvania and the Linux Foundation to create an open-source data science curriculum kit that any institution can use.

"There's tremendous need in the market for building something that will accelerate and will help academic institutions build their own programs," Ana Echeverri, IBM's AI Skills Learning and Certifications Lead, told Education Dive.

The kit, which is expected to be available on GitHub early next year, will be composed of independent modules that teach the core data science concepts. Each one will include slides, lecture notebooks and homework that instructors can use as is or alter to fit their needs.

U of Pennsylvania researchers are creating the curriculum by drawing from one of its own courses that teaches students the broad principles of data science, Davidson said.

Once those training materials are available, other universities will also be able to contribute their own open-source content to the project. An oversight committee will approve changes to existing material or new additions.

The aim, those working on the project say, is to help colleges get data science programs off the ground faster and to align them with industry needs. "It's great to have the view of, 'What are the skills that industry is actually expecting our people to have in order to do this job?'" Echeverri said.

U of Pennsylvania is helping create open-source curriculum for data science programs. | Credit: Bryan Y.W. Shin/Wikimedia Commons

A field 'in its infancy'

Creating a data science program from scratch can be tricky.

For one, it's still an emerging field. "Despite the expansion of data science activities that's really been driven by industry and now is being embraced by academia, ... this field is in its infancy," said Julia Lane, executive director of the University of Chicago's Center for Data and Computing. The center, which opened last year, aims to help define the field, she added.

And it can be hard to pick what a program will cover because data science draws on many disciplines — including computer science, mathematics and statistics — and has applications in almost every field. For example, delivery companies may use data science to optimize when they drop off packages, while the health care sector may use it to improve medical imaging.

Moreover, the nationwide shortage of data scientists is a relatively recent problem, giving colleges little time to adapt. Just four years ago, there were more workers with data science skills than there were jobs that required them, LinkedIn notes.

From 2015 to 2018, job postings for data scientists on Indeed.com rose by 75%. That trend continued in 2019, with a 56% year-over-year increase in data science job openings, according to an annual report from LinkedIn.

Industry leaders and alternative education providers, which often can move more quickly than higher ed institutions, have stepped in with their own offerings. Last year, IBM rolled out a data science professional certificate on Coursera. And just this month, Coding Dojo announced a 14-week boot camp to train students in data science.

Building an interdisciplinary program

While many colleges have been scrambling to keep up, some have been making major investments into their data science offerings to get ahead.

The Massachusetts Institute of Technology (MIT), for instance, announced last year that it was investing $1 billion into a new computing college that will bring together several disciplines, including computer science, data science and the humanities.

Just across the river from MIT, Boston University is making a statement by erecting a 19-story building to house its new data science center, though the design has received some pushback. "We wanted architecture that would signal to everyone that this was a special place, the center of campus," its president, Robert Brown, said in the announcement last year.

> **"Data science is an incredibly hot area. Yet there are relatively few places and resources that currently are available for teaching data science."**
>
> Susan Davidson
> *Computer and information science professor, University of Pennsylvania*

Others are taking a different approach. With the help of a $120 million donation, the University of Virginia (UVA) is launching an interdisciplinary school of data science with the goal of breaking down academic silos. To do so, the university is setting up satellites, which it is calling "collaboratories," across campus, with the first housed inside its school of education.

"We want to make sure we're completely connected with other programs," Donald Brown, founding director of UVA's School of Data Science, told Education Dive in an interview. He expects there will eventually be data science "collaboratories" in each of the university's schools.

So far, UVA only offers master's degrees in data science. Standing up an undergraduate program can be more difficult, Brown said, due to the need to strike a balance between teaching students technical skills required in the workplace and providing the

general education courses that will help them be well-rounded adults.

"We don't want to have a narrowly focused program," he said.

But IBM's Echeverri predicts data science skills will eventually be taught to all undergraduates. "Even if there are separate data science programs," she said, "the reality is that data science is going to become embedded in a lot of (other) programs that are not defined today as technical programs."