Harvard Data Science Review • Issue 3.1, Winter 2021

Is Data Science Education a Jack of All Trades?

David R. Hardoon¹

¹University College London, United Kingdom

Published on: Feb 25, 2021

DOI: <https://doi.org/10.1162/99608f92.d1aaf1d7>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

The demand for data science professionals has been echoed in the increased supply of data science education. In the field of data science alone, we have witnessed a flourishing of formal degrees, diplomas, and certifications, among other forms of training. However, has our eagerness in producing data scientists resulted in the proliferation of generalists rather than practical experts in the field?

What is data science? Unfortunately, this is a complex question with a diverse answer depending on who you ask. Interestingly, it is the diversity of this answer that results in the quandary as to the skills required to be acquired by a data science professional.

According to the *Wikipedia* (“Data Science,” 2021) definition, in the broad sense data science “is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from many structural and unstructured data.” The field is a hybrid of statistics, machine learning, and data mining, and covers programming languages, technology frameworks, development platforms, as well as visualization tools. In other words, data science is the umbrella under which all these individual and/or intersecting disciplines sit.

The diversification and complexity of data science poses a real academic challenge in data science education. How does one balance the multiplicity of areas, covering them with sufficient depth in order to produce a well-rounded data scientist who can provide impactful value to a future employer? The article by Eric Kolaczyk, Haviland Wright, and Masanao Yajima (Kolaczyk et al., 2020) highlighted such a challenge where existing degree programs “can leave students, upon exiting academia, needing a nontrivial ramping-up period before they can truly have an impact with their first employers.” This feedback is common and goes beyond that of the applicability of the practicum component of a degree. The majority of data scientists, upon completion of their education, still require a ‘hands-on’ education at their place of employment to refine their learned skills to practical necessities. In other words, a ‘fresh’ well-rounded data scientist is unlikely to hit the ground running at their first employ despite a presumption otherwise.

How then should well-rounded data scientists be produced—so that they can be effective, as close as possible to day one, upon exiting their respective training? Rather than trying to cover the breadth of all that is data science, academia may wish to recognize that the field has matured and evolved in its pedagogy taxonomy.

Focusing on professional application, rather than pure academic research, we can identify the following four key example roles that an organization may need in operationalizing their data science ambitions.

1. The data science developer (or modeler), who would be the core developer with a stronger focus on data science methodologies and techniques.
2. The data science engineer, who would focus on data piping, data quality, and data ingestion for the purpose of data science modeling.

3. The data science solution architect, who would have a deeper understanding of platforms and data enterprise architecture for an end-to-end operationalization.
4. The data science storyteller, who would have a stronger focus on data visualization and data science communication and possess strong business process acumen.

The aforementioned roles are not just an application of data science. While having overlapping areas, each carries an in-depth focus, specific required competencies, skillsets, and practicum considerations. These elements highlight that a unified data science program may not do any of the respective areas justice (Davenport, 2020). Worse, an employer may hire using the generic description of data scientist without realizing they may need one specific area over the others, or indeed require four separate hires rather than the one.

Whether through the creation of dedicated degrees with specific focus areas, diversification of modules that allow the flexibility of commonality and a major/minor approach, or a degree that differentiates between a research and a practicum (professional) focus, it is clear that there is a need to refocus our efforts. Data science academic programs should tailor data science education to produce professional competency with sufficient depth that enables immediate demonstrable outcome upon completion. Such programs need to reflect the maturity of the data science field and allow more specificity in the professional expertise they aspire to produce. Alternatively, or perhaps concurrently, employers need to appreciate that newly minted data scientists are not quite ‘road-ready’ and may require secondary on-the-ground training in order to contextualize and surface the benefits anticipated from data scientists.

Disclosure Statement

David R. Hardoon has no financial or non-financial disclosures to share for this article.

References

- Davenport, T. (2020). Beyond unicorns: Educating, classifying, and certifying business data scientists. *Harvard Data Science Review*, 2(2). <https://doi.org/10.1162/99608f92.55546b4a>
- Kolaczyk, E., Wright, H., & Yajima, M. (2020). Statistics practicum: Placing “practice” at the center of data science education. *Harvard Data Science Review*, 3(1). <https://doi.org/10.1162/99608f92.2d65fc70>
- Data science. (2021, January 19). In *Wikipedia*. https://en.wikipedia.org/wiki/Data_science
-

©2021 David R. Hardoon. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.