

Deployment as a Critical Business Data Science Discipline

Thomas Davenport^{1,2} Katie Malone³

¹Division of Operations and Information Management, Babson College, Babson Park, Massachusetts, United States of America,

²International Institute for Analytics, Portland, Oregon, United States of America,

³Tempus Labs, Chicago, Illinois, United States of America

Published on: Feb 10, 2021

DOI: <https://doi.org/10.1162/99608f92.90814c32>

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

Column Editors' Note: *In this article, we focus on a key problem in industry: getting data science models deployed into production within organizations. The tasks and skills involved in deployment are often not considered as a key component of data science initiatives, but they are critical to data science success. We describe evidence of the deployment problem, the components of deployment, and how some campus-based business analytics degree programs attempt to inculcate deployment skills.*

Keywords: data science, deployment, business analytics, change management

It is becoming increasingly clear that deployment—getting analytical and artificial intelligence (AI) systems fully and successfully implemented within organizations—is becoming one of the most critical disciplines at all phases of a business data science project. We often incorrectly think of deployment of a data science or analytical model as the last stage of the process, when the model or algorithm-based system is put into production as a part of a business process. A key aspect of deployment is change in the business process: a successfully deployed model will take a set of tasks that had previously been manual, heuristics-based, or simply impossible, and insert an algorithmically based solution. Starting with the algorithm first, and only at the end of the project thinking about how to insert it into the business process, is where many deployments fail. Instead of thinking of deployment as the last step in a linear set of activities, a data scientist—or at least key members of data science teams—should consider factors that have a strong influence on deployment throughout the data science project.

For example, in the large-scale UPS Orion project to provide real-time optimized routing for the company's truck drivers, a majority of the 10-year implementation cycle and the several-hundred-million-dollar budget was spent on deployment and change management issues (Center for Technology and Sustainability, 2016). The complex algorithm needed to be implemented on servers and handheld devices, drivers and their supervisors needed to be convinced that the algorithm worked effectively, and multiple package sorting and loading processes needed to be changed. UPS isn't unique: many firms with effective data-driven adaptations have needed to overcome similar deployment hurdles in order to realize gains. However, deployment skills like stakeholder analysis and change management are comparatively undercovered in traditional data science training programs, leaving many data scientists underprepared for this difficult part of the process.

One implication of the deployment challenge is that data scientists need to anticipate and be able to respond to deployment issues. Yet many data scientists are not trained in how to identify and address deployment issues, and may not view deployment as a part of their jobs. Anecdotally, when we have asked data scientists or data science leaders what percentage of their models are deployed, few of them know (Davenport, 2019).

Some organizations leave data scientists to focus on creating models and coding, while leaving organizational or technical deployment issues to other roles. At Southern California Edison, for example, a data scientist is

paired with a “Predictive Analytics Advisor” who handles relationships with the client department and is responsible for organizational change issues (Davenport, 2020b). These roles are also called ‘product managers,’ ‘analytics translators,’ or ‘data/analytics/AI strategists’ (Henke et al., 2018). Other companies stress ‘data engineering’ or ‘machine learning engineering’ roles in handling technical deployment issues. One large bank that we worked with referred to these roles as ‘light quants’—a characterization that may be technically correct but is disparaging of their importance to successful projects.

It is increasingly recognized that data scientists can’t be ‘unicorns’ and possess all needed types of expertise on projects (Davenport, 2020a). Therefore, it is often desirable to assign primary deployment responsibilities to someone on a data science team who is trained and evaluated on the overall successful implementation of the project. At a minimum, however, even model and coding-oriented data scientists need to be aware of deployment issues and bring them to the attention of specialists if they exist. If their models aren’t deployed, both the models and the data scientists who create them may be of little value to employers. The entire domain of data science may lose favor within an organization if models are only rarely deployed. And for those industries where auditability and transparency are absolutely critical, such as banking, finance, and health care, a poorly deployed model is a legal, business, or health risk.

Evidence of a Deployment Problem

Surveys of organizations and market research reports in the United States and globally suggest that deployment challenges are widespread. Rexer Analytics surveys of data scientists over several years found that only a small fraction of respondents say that all their models are deployed. Karl Rexer, who conducts the surveys, commented: “Predictive model deployment continues to be a major challenge for many companies...In our 2017 Data Science Survey, only 13% of data scientists say their models always get deployed. And deployment is not improving: with each survey, going back to 2009 when we first asked this question, we see almost identical results” (Alteryx, 2017). In one of the surveys, companies in which models were more likely to be deployed had data scientists with substantially higher job satisfaction (Allen et al., 2015).

A survey of large financial services and life sciences firms in 2019 by NewVantage Partners found that firms were actively embracing AI technologies and solutions, with 91.5% of firms reporting ongoing investment in AI. But only 14.6% of firms reported that they have deployed AI capabilities into widespread production (NewVantage Partners, 2019).

In a 2019 global McKinsey survey with the headline “AI Adoption Proves Its Worth, But Few Scale Impact,” between 12% (in consumer packaged goods) and 54% (in high-tech firms) of respondents had at least one machine learning application implemented in a process or product. Only 30% of respondents overall were using AI in products or processes across multiple business units and functions (Cam et. al, 2019).

In a 2020 [MIT Sloan Management Review/BCG survey](#) of global executives, only one in 10 companies reported significant financial benefits from implementing AI in their companies (Ransbotham et al., 2020). The

survey report authors emphasize the importance of deployment: “But it’s the final stage of AI maturity, of successfully orchestrating the macro and micro interactions between humans and machines, that really unlocks value.”

Market research firms have also concluded that many organizations face difficult deployment issues. For example, a Gartner report on the future of AI commented, “the reality is that most organizations struggle to scale the AI pilots into enterprise wide production, which limits the ability to realize AI’s potential business value” (Costello, 2020).

Another market researcher, Forrester, suggests that a broad range of skills is necessary for successful business (not just technical) AI deployment:

AI is by nature a learning system, which forces a continuous development cycle to advance AI’s business contribution. AI projects require broad expertise beyond tech development and deployment due to the need for business change management. Building trust in automated decisions is essential to arrive at an AI that is accepted by customers and employees. (Granzen, 2020)

These sources provide support for the idea that deployment is an important but largely underaddressed aspect of data science. However, in order to address the issue of deployment, we need a better understanding of what it is.

What Are the Components of Deployment?

A critical early step is understanding the business problem and environment to which the model will be applied. This understanding typically requires the data scientist to observe in detail the current process and people performing it. It includes assessing the current decision-making approach, understanding what data might be available and how they are collected (whether it is collected from humans or machines), and ascertaining the skills and capabilities of the people who do the work today (Birnbaum, 2004; Horrell et al., 2020). Embedding the output in a product requires understanding the customer’s needs and how existing products do or do not fulfill them. For traditional products this is often the realm of a product manager. Even when data science teams also have a product management role, having the technical expert (i.e., the data scientist) in direct contact with users is important for establishing context for how the model will eventually be used.

At this early stage of data science, the data scientist is gaining critical insight into what are the types of mistakes that the future model can make and the implications of each one, which will help him or her tune the model during its development. This is also the stage at which a thoughtful data scientist will start thinking about what aspects of the model will need to be monitored when it goes into deployment, and how to surface those metrics to the right stakeholders.

Another important precursor of deployment is gaining the trust of the manager or other stakeholders responsible for the process or product. Those individuals will make the decision whether or not to deploy the new model, and the technical details of how it works may never be accessible to the stakeholders. Therefore, it is important that they trust that the data scientist understands the problem and has a handle on how to solve it and that the model will not introduce new complications into the business domain.

Before building the model, the data scientist also needs to be familiar with the details of the technology stack into which the model will fit. A highly deployable model will be tightly integrated with existing systems, will not require a lot of new skills or tasks on the part of the user, and will not require any major changes in the technology architecture. Those that do require major changes in technology will, of course, take longer, cost more, and be more difficult to deploy. When models require heavy technical adaptation to deploy, the data scientist should extend his or her research and trust-building phase to include IT personnel.

Critically, ‘infrastructure’ at this point also includes the data sources and the pipelines through which data flows from its source to the algorithm. One failure mode for models that seemed perfectly performant during development is that, come deployment time, the data being fed into the algorithms is not reliable, doesn’t update regularly, changes unpredictably, or otherwise has some difference from the data that was used to develop the model. Someone on the analytics or AI project team must ensure that data is accessible, of sufficient quality, and in large enough volumes to support model development and deployment. Trust that the data or model don’t contain bias and have an ethical purpose is also required. The ‘AI data czar’ is a role that lacks a consensus title and is forgotten by many organizations, so it is often folded into the role of the data scientist (Beck et al., 2019).

Insofar as deployment issues can be monitored and managed with technology, that technology should be considered (some data science tools and platforms, like Domino Data Lab, DataRobot, Databricks, and H2O.ai have features that support or automate aspects of deployment, and are worth considering at this stage even if they weren’t part of the model development process). For the many problems that a data scientist cannot anticipate at this stage, because production data systems tend to find new and interesting ways of breaking all the time, having a human checking in on the model is often a good idea. Notably, the significant technical expertise required for this aspect of deployment leads to another division of labor; the ‘machine learning engineer’ role is increasingly employed at this point in the process.

As the model is being built and refined, the data scientist needs to address not only the fit of the model to the data, but also accountability for changes it creates in the outcome of the process, both immediately and in the future. The choices that the data scientist makes at this stage can have a significant impact on how easy those future updates will be to make. If a complex and nonexplainable model fits the data better than an explainable one, is the increased model fit worth a reduction in likely deployability? If new features require new data, will frontline users be able to provide it? If the model requires a GPU-based processor to run, will that work with the existing systems architecture?

Finally, after the model has been developed and a program (or application program interface--API) created for the deployment of the solution, there are still tasks to perform. A data scientist or deployment-focused team member should be monitoring how the model and related systems perform as they are first operated. Are the users employing the model and system appropriately, and making the right decisions? Are the early cases providing economic value? Does the model give consistent and replicable results, especially as the underlying data set updates? Are the model's predictions presenting any unforeseen ethical issues, or displaying bias? Some data science team member at this point should be playing the role of an economist, providing insights on whether the new model is achieving return on investment. However, many data scientists who have not been trained in this 'maintenance mode' phase of model development usually have moved on to building the next model at this point, leaving a highly technical and often difficult-to-debug system behind for someone else to deal with.

It should be apparent, then, that many of these factors driving a successful deployment are not typically thought of or taught as data science. Yet if these factors are not well-managed in a process of data science implementation, no overall value is likely to be realized. Any process that helps data scientists or analysts to carry out better and more relevant data analysis should be an integrated part of data science. Those organizations that don't have the data science infrastructure (technical but also organizational and strategic) that would allow a data scientist to specialize in model-building should either be creating such an infrastructure, or incorporating deployment into the job description of data scientists. If they do neither they are condemning model-focused data scientists to have a very difficult time having an impact.

Teaching and Developing Deployment Skills

What are some options for making sure data scientists are well-prepared to deploy models on the job? Traditional education has (so far) shown some shortcomings, leaving employers and data scientists largely to rely on themselves. But as the field matures, savvy educators who are aware of the 'deployment gap' can address it in curricula.

The American higher education system is generally quite successful at teaching many of the other skills expected of data scientists, such as computer programming, statistics, and machine learning. However, since many of the deployment capabilities are contextual and relationship based, it is difficult to teach them in a traditional instructor-led course. The only sure way to develop them is to create experiences in which deployment of a data science project takes place. Even there it may be difficult to establish accountability for models, decisions, and actions developed in courses or capstone projects.

Advanced degree programs in data science face challenges to incorporating such deployment experiences. These programs are often found in computer science and engineering schools, which may not have relationships with businesses that sponsor project work, and do not often include organizational change or other deployment topics in their curricula. Business schools do often have those relationships and typically

require a class or two in organizational behavior and change, but some faculty may not be convinced of the need for hands-on projects if they come at the expense of more technical instruction. In addition, many master's degree programs in data science are exclusively online—a medium that is not amenable to many real-life experiences. Ph.D. programs that lead to data science careers—from physics to statistics to anthropology—are not likely to be focused on business deployment issues, even within a single course.

For teaching technical aspects of deployment, one promising option is campus-based master's programs in business analytics that are more likely to include deployment-oriented content and a project that involves deployment issues. One of the earliest of these programs (beginning in 2007), the [Master of Science in Analytics](https://analytics.ncsu.edu/) (<https://analytics.ncsu.edu/>) at North Carolina State, has a team-based practicum project with businesses, government agencies, and nonprofits that takes place over 8 months. Michael Rappa, the program's director, commented in an email (June 11, 2020) that:

Students learn that at the heart of a data scientist's work is a process: understanding the business need, devising an analytical framework, wrangling the data, building models, drawing insights, communicating results, and should the results warrant it, packaging a deliverable that can be deployed in a production environment. It's all about process. If the team upholds the integrity of the process, the end result will be useful—even if it's not used. If the team circumvents the process or undermines its integrity, then whatever it delivers is useless—no matter how promising it may seem you cannot be confident in the results... Students learn to respect the rigor of the process grappling with the many ways it can go sideways when you're not disciplined. It can be frustrating for students weaned on textbook problems, and humbling. The lessons learned stick with them.

For understanding how deployment fits into the end-to-end modeling process, doing projects in collaboration with industry gives students valuable real-world experience. For example, at MIT, the relatively new [Master of Business Analytics](#) program has an Analytics Lab with company projects, a Capstone Project over 7 months, and a course that addresses deployment-related skills. The three courses are described as follows:

Analytics Lab: Matches student teams with leading-edge projects involving analytics, machine learning, or digital technologies as they apply to business questions and problems.

Capstone Project: This 7-month project pairs two students with organizations proposing an analytics project, and normally includes both remote work for a semester and a 10-week summer residence period on the project site.

From Analytics to Action: Develops appreciation for organizational dynamics and competence in navigating social networks, working in a team, demystifying rewards and incentives, understanding change initiatives, and making sound decisions.

Michelle Li, the director of the Business Analytics program, wrote in an email (June 22, 2020) about the purpose of these courses:

In these courses, students must work in teams on hands-on analytics problems with partner organizations. In doing so, students are exposed to not only messy, disparate and incomplete datasets but also organizational and cultural barriers which can make academic modeling techniques difficult to implement in the real world. Once the Master's students complete the 1-year immersion program, they are more prepared to tackle large data science problems which require both technical and deployment-oriented skills such as business communication, understanding of organizational behavior and a focus on relationship-building.

Data scientists weighing their educational options should examine the project experiences they will receive at each institution, much like they might consider the name recognition of the school and the in-class curricula.

Of course, there are ways that students can acquire deployment expertise and experience outside of a formal educational program. Unless project work and deployment experience become the norm for data science education, employers and data scientists should assume that much of the burden is on them to teach and learn this important skillset. Previous jobs, for example, may have provided a given data scientist with on-the-job lessons in deployment. Even if the job experience does not involve data science or analytics, other types of projects can certainly foster the deployment-oriented skills and sensibilities. One suggestion for data science managers is to have newly hired junior data scientists start with a 'rotation' maintaining or expanding a model in production, as a form of service to the organization and a training ground. Seeing the end state early on can produce a very different frame of mind than starting out by reading papers and creating models—a more common default for new data science hires.

Organizational Steps to Improve Deployment Success

There are several steps that organizations can take to improve the likelihood that projects are successfully deployed. They include:

- Creating data science teams whose members specialize in certain skills, but all contribute to deployment. The teams may include model-oriented data scientists, machine learning engineers, product managers, data engineers, and technical operations members, and they should jointly own deployment and each contribute to it in some way.
- When hiring data scientists and data science team members, asking about experience with deployment in job interviews. One chief analytics officer we know abruptly ends interviews if the candidate cannot recall and describe a successful deployment in previous jobs.
- Making clear the deployment-related responsibilities of data scientists and data science team members in job definitions and project assignments.

- Establishing a ‘pipeline’ toward deployment for each data science project that culminates in production deployment. As with ‘stage gate’ product development processes or lean product development efforts beginning with minimum viable products, there may be attrition along the way, but the goal is to fully deploy. Farmers Insurance is one company that has established such a pipeline (Davenport and Bean, 2018). And in a 2018 Deloitte survey, 54% of the U.S. executive respondents in large firms working with AI said they already had a process for moving prototypes into production (Deloitte Insights, 2018).
- Create ‘air cover’ for successful deployments by cultivating buy-in for projects from senior management. At a large pharmaceutical firm, for example, the chief data and analytics officer got consensus from the senior management team for three major projects, each with considerable value to the firm. The commitment included sufficient human and financial resources for taking the projects into full deployment, assuming that interim objectives were met.
- If an organization has a data-driven culture—characterized by good practices around incentivizing the use of data, encouraging data literacy among employees, and supporting data-driven decision making in the executive ranks—a given data science initiative has much more fertile ground for gaining adoption.

Of course, some data scientists employed in businesses may simply be uninterested in acquiring deployment-oriented skills, perhaps feeling that their only role is to create effective and innovative algorithms. But in the end, the great majority of business data scientists’ jobs involve solving business problems with data and analysis, and the algorithm is only a means to that end. If algorithm creation alone is their orientation, however, they should be aware that a model or algorithm that is not deployed—no matter how technically brilliant—means that the job has not been finished and that their organizations will get little value from their efforts. At a minimum, data scientists need to respect the need for deployment-oriented skills and the people who bring them to projects.

Acknowledgments

We are grateful for the comments of three anonymous reviewers and the Editor-in-Chief of *HDSR*.

Disclosure Statement

Thomas Davenport is an advisor to three companies mentioned in this column: NewVantage Partners, DataRobot, and Deloitte. He is also a research fellow at MIT, but is not affiliated with the MIT Master of Business Analytics program.

References

Allen, H., Gearan, P., & Rexer, K. (2015). *2015 data science survey*. Rexer Analytics.
www2.cs.uh.edu/~ceick/UDM/Rexer2015.pdf

- Alteryx. (2017, September 12). *Alteryx Promote announced: New offering to easily deploy, manage and integrate data science models for real-time decisions* [Press release]. <https://www.alteryx.com/press-releases/2017-09-12-alteryx-promote-announced-new-offering-easily-deploy-manage-and-integrate>
- Beck, M., Davenport, T., & Libert, B. (2019, March 14). The AI roles some companies forget to fill. *Harvard Business Review*. <https://www.hbr.org/2019/03/the-ai-roles-some-companies-forget-to-fill>
- Birnbaum, M. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, 55, 803–832. <https://www.doi.org/10.1146/annurev.psych.55.090902.141601>
- Cam, A., Chui, M., & Hall, B. (2019, November 22). *Global AI Survey: AI proves its worth, but few scale impact*. McKinsey & Company. <https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact>
- Center for Technology and Sustainability. (2016, March 30). *Looking under the hood: ORION technology adoption at UPS*. Business for Social Responsibility. <https://www.bsr.org/en/our-insights/case-study-view/center-for-technology-and-sustainability-orion-technology-ups>
- Costello, K. (2020, Feb. 5). *Gartner predicts the future of AI technologies*. Gartner. <https://www.gartner.com/smarterwithgartner/gartner-predicts-the-future-of-ai-technologies/>
- Davenport, T. (2019, March 21). What's your deployment score? *International Institute for Analytics*. <https://www.iianalytics.com/blog/2019/3/21/whats-your-deployment-score>
- Davenport, T. (2020a, May 19). Beyond unicorns: Educating, classifying, and certifying business data scientists. *Harvard Data Science Review*, 2(2). <https://doi.org/10.1162/99608f92.55546b4a>
- Davenport, T. (2020b, July 30). Machine learning and organizational change at Southern California Edison. *Forbes*. <https://www.forbes.com/sites/tomdavenport/2020/07/30/machine-learning-and-organizational-change-at-southern-california-edison/#1767a96f3336>
- Davenport, T., & Bean, R. (2018, August 1). Farmers accelerates its time to impact with AI. *Forbes*. <https://www.forbes.com/sites/tomdavenport/2018/08/01/farmers-accelerates-its-time-to-impact-with-ai/?sh=56308d4b672a>
- Deloitte Insights. (2018). *State of AI in the enterprise* (2nd ed.). https://www2.deloitte.com/content/dam/insights/us/articles/4780_State-of-AI-in-the-enterprise/DI_State-of-AI-in-the-enterprise-2nd-ed.pdf
- Granzen, A. (2020, February 17). Consultancies are reinventing their service model for AI. *Forrester*. <https://go.forrester.com/blogs/consultancies-are-reinventing-their-service-model-for-ai/>

Henke, N., Levine, J., & McInerney, P. (2018, February 5). You don't have to be a data scientist to fill this must-have analytics role. *Harvard Business Review*. <https://hbr.org/2018/02/you-dont-have-to-be-a-data-scientist-to-fill-this-must-have-analytics-role>

Horrell, M., McElhinney, A., & Reynolds, L. (2020, April 30). Data science in heavy industry and the Internet of Things. *Harvard Data Science Review* 2(2). <https://doi.org/10.1162/99608f92.834c6595>

NewVantage Partners. (2019). *Big Data and AI Executive Survey of 2019*. <https://www.newvantage.com/wp-content/uploads/2018/12/Big-Data-Executive-Survey-2019-Findings-Updated-010219-1.pdf>

Ransbotham, S., Khodabandeh, S., Kiron, D., Candelon, F., Chu, M., & Lafountain, B. (2020, October 19). Expanding AI's impact with organizational learning. *MIT Sloan Management Review*. https://sloanreview.mit.edu/projects/expanding-ais-impact-with-organizational-learning/?utm_medium=pr&utm_source=release&utm_campaign=ReportBCGAI2020

©2021 Thomas Davenport and Katie Malone. 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.

Spanish translation provided by IAE Business School, Austral University, Argentina.



[Davenport & Malone \(2021\) Spanish Version.docx](#)

37 KB