

FEATURE

Essential skills and traits of elite data scientists

These days the data scientist is king. But extracting true business value from data requires a unique combination of technical skills, mathematical know-how, storytelling, and intuition.

By Bob Violino

Contributing Writer, CIO

MAR 27, 2018 3:00 AM PDT

Data scientists continue to be in high demand, with companies in virtually every industry looking to get the most value from their burgeoning information resources.

“As organizations begin to fully capitalize on the use of their internal data assets and examine the integration of hundreds of third-party data sources, the role of the data scientist will continue to expand in relevance,” says Greg Boyd, director at consulting firm Protiviti.

“In the past, the teams responsible for data were relegated to the back rooms of the IT organization, performing the critical database tasks to keep the various corporate systems fed with the data ‘fuel’ [that] allowed corporate executives to report out on operations activities and deliver financial results,” Boyd says.



BrandPost Sponsored by HPE

What Is Consumption-Based IT and Is It Right For Your Business?

This role is important, but the rising stars of the business are those savvy data scientists that have the ability to not only manipulate vast amounts of data with sophisticated statistical and visualization techniques, but have a solid acumen from which they can derive forward-looking

insights, Boyd says. These insights help predict potential outcomes and mitigate potential threats to the business.

[Thinking of making a shift to data science? Find out all you need to know about the lucrative data scientist role. | Prove your data science chops by earning one of these data science certifications. | Get the latest analytics insights by signing up for our CIO Leader newsletter.]

So what does it take to be data science whiz? Here are some important attributes and skills, according to IT leaders, industry analysts, data scientists, and others.

Critical thinking

Data scientists need to be critical thinkers, to be able to apply objective analysis of facts on a given topic or problem before formulating opinions or rendering judgments.

“They need to understand the business problem or decision being made and be able to 'model' or 'abstract' what is critical to solving the problem, versus what is extraneous and can be ignored,” says Anand Rao, global artificial intelligence and innovation lead for data and analytics at consulting firm PwC. “This skill more than anything else determines the success of a data scientist,” Rao says.

A data scientist needs to have experience but also have the ability to suspend belief, adds Jeffry Nimeroff, CIO at Zeta Global, which provides a cloud-based marketing platform.

“This trait captures the idea of knowing what to expect when working in any area, but also knowing that experience and intuition are imperfect,” Nimeroff says. “Experience provides benefits but is not without risk if we get too complacent. This is where the suspense of belief is important.”

It’s not about looking at things with the wide eyes of a novice, Nimeroff says, but instead stepping back and being able to assess a problem or situation from multiple points of view.

Coding

Top-notch data scientists know how to write code and are comfortable handling a variety of programming tasks.

“The language of choice in data science is moving towards Python, with a substantial following for R as well,” Rao says. In addition, there are a number of other languages in use such as Scala, Clojure, Java and Octave.

“To be really successful as a data scientist, the programming skills need to comprise both computational aspects — dealing with large volumes of data, working with real-time data, cloud computing, unstructured data, as well as statistical aspects — [and] working with statistical models like regression, optimization, clustering, decision trees, random forests, etc.,” Rao says.

The impact of big data beginning in the late 1990s has demanded that more and more data scientists understand and be able to code in languages such as Python, C++ or Java, says Celeste Fralick, chief data scientist at security software company McAfee.

If a data scientist doesn’t understand how to code, it helps to be surrounded by people who do. “Teaming a developer with a data scientist can prove to be very fruitful,” Fralick says.

Math

Data science is probably not a good career choice for people who don’t like or are not proficient at mathematics.

“In our work with global organizations, we engage with clients looking to develop complex financial or operational models,” Boyd says. “In order for these models to be statistically relevant, large volumes of data are required. The role of data scientist is to leverage their deep expertise in mathematics to develop statistical models which may be used to develop or shift key business strategies.”

The data scientist whiz is one who excels at mathematics and statistics, while having an ability to collaborate closely with line-of-business executives to communicate what is actually happening in the “black box” of complex equations in a manner that provides re-assurance that the business can trust the outcomes and recommendations, Boyd says.

Machine learning, deep learning, AI

Industries are moving extremely fast in these areas because of increased compute power, connectivity, and the huge volumes of data being collected, Fralick says. “A data scientist needs to stay in front of the curve in research, as well as understand what technology to apply when,” she says. “Too many times a data scientist will apply something ‘sexy’ and new, when the actual problem they are solving is much less complex.”

Data scientists need to have a deep understanding of the problem to be solved, and the data itself will speak to what’s needed, Fralick says. “Being aware of the computational cost to the ecosystem, interpretability, latency, bandwidth, and other system boundary conditions — as well as the maturity of the customer — itself helps the data scientist understand what technology to apply,” she says. That’s true as long as they understand the technology.

Also valuable are statistical skills. Most employers do not consider these skills, Fralick says, because today’s automated tools and open source software are so readily available. “However, understanding statistics is a critical competency to comprehending the assumptions these tools and software make,” she says.

It’s not enough to understand the functional interfaces to the machine learning algorithms, says Trevor Schulze, CIO at data storage provider Micron Technology. “To select the appropriate algorithm for the job, a successful data scientist needs to understand the statistics within the methods and the proper data preparation techniques to maximize overall performance of any model,” he says.

Skills in computer science are also important, Schulze says. Because data science is mainly done at the keyboard, strong fundamentals in software engineering are helpful.

Communication

The importance of communication skills bears repeating. Virtually nothing in technology today is performed in a vacuum; there’s always some integration between systems, applications, data and people. Data science is no different, and being able to communicate with multiple stakeholders using data is a key attribute.

“The ‘storytelling’ ability through data translates what is a mathematical result into an actionable insight or intervention,” says Rao. “Being at the intersection of business, technology, and data, data scientists need to be adept at telling a story to each of the stakeholders.”

That includes communicating about the business benefits of data to business executives; about technology and computational resources; about the challenges with data quality, privacy, and confidentiality; and about other areas of interest to the organization.

Being a good communicator includes the ability to distill challenging technical information into a form that is complete, accurate, and easy to present, Nimeroff says. “A data scientist must remember that their execution yields results that can and will be used to support directional action by the business,” he says. “So, being able to ensure that the audience understands and appreciates everything that is being presented to them — including the problem, the data, the success criteria, and the results — is paramount.”

A good data scientist must have the business savvy and inquisitiveness to adequately interview the business stakeholders to understand the problem and identify which data is likely to be relevant, Schulze says.

In addition, data scientists need to be able to explain algorithms to business leaders. “Communicating how an algorithm arrived at a prediction is a critical skill to gain leaders’ trust in predictive models being part of their business processes,” Schulze says.

Data architecture

It is imperative that the data scientist understand what is happening to the data from inception to model to business decision.

“To not understand the architecture can have serious impact on sample size inferences and assumptions, often leading to incorrect results and decisions,” Fralick says.

Even worse, things can change within the architecture. Without understanding its impact on models to begin with, a data scientist might end up “on a firestorm of model redo’s or suddenly inaccurate models without understanding why,” Fralick says.

While Hadoop gave big data legs by delivering the code to the data and not vice versa, Fralick says, understanding the complexities of the data flow or data pipeline are critical to insuring good fact-based decision-making.

Risk analysis, process improvement, systems engineering

A sharp data scientist needs to understand the concepts of analyzing business risk, making improvements in processes, and how systems engineering works.

“I’ve never known an excellent data scientist without these” skills, Fralick says. “They all play hand-in-hand, both inwardly focused to the data scientist but outwardly to the customer.”

Inwardly, the data scientist should remember the second half of the title — scientist — and follow good scientific theory, Fralick says.

Building in risk analyses at the start of model development can mitigate risks. “Outwardly, these are all skills that data scientists require to probe the customer about what problem they are trying to solve,” she says.

Connecting spending to process improvement, comprehending inherent company risks and other systems that can impact data or the result of a model can lead to greater customer satisfaction with the data scientist’s efforts, Fralick says.

Problem solving and good business intuition

In general, the traits great data scientists exhibit are the same traits that are exhibited by any good problem solver, Nimeroff says. “They look at the world from many perspectives, they look to understand what they are supposed to be doing before pulling all the tools out of their tool belt, they work in a rigorous and complete manner, and they can smoothly explain the results of their execution,” Nimeroff says.

When evaluating technology professionals for roles such as data scientists, Nimeroff looks for these traits. “The approach yields far more successes than failures, and also ensures that potential upside is maximized because critical thinking is brought to the forefront.”

Finding a great data scientist involves finding someone who has somewhat contradictory skill sets: intelligence to handle data processing and create useful models; and an intuitive understanding of the business problem they’re trying to solve, the structure and nuances of the data, and how the

models work, says Lee Barnes, head of Paytronix Data Insights at business software provider Paytronix Systems.

“The first of these is the easiest to find; most people with good math skills and a degree in math, statistics, engineering, or other science-based subjects are likely to have the intellectual horsepower to do it, Barnes says. “The second is much harder to find. It is surprising how many people we interview that have built complex models, but when pushed on why they think the model worked or why they chose the approach they did, they don’t have a good answer.”

These people are likely to be able to explain how accurate a model was, “but without understanding why and how it works, it’s hard to have a lot of confidence in their models,” Barnes says. “Someone with this deeper understanding and intuition for what they are doing is a true data science whiz, and will likely have a successful career in this field.”

Related data science and data analytics articles:

Page 1 of 2



 **Survey says! Share your insights in our 2020 CIO Tech Poll.**



Copyright © 2019 IDG Communications, Inc.