Harvard
Business
Review

Hiring

The Question to Ask Before Hiring a Data Scientist

by Michael Li

August 06, 2014

When hiring data scientists, there's nothing more frustrating than making the wrong hire. Data scientists are in notoriously high demand, hard to attract, and command large salaries — compounding the cost of a mistake. At The Data Incubator, we've talked to dozens of employers looking to hire data scientists from our training program, from large corporates like Pfizer and JPMorgan Chase to smaller tech startups like Foursquare and Upstart. Employers that didn't have good hiring experiences in the past often failed to ask a key question:

> *Is your data scientist producing analytics for machines or humans*?

This distinction is important across organizations, industries, and job titles (our fellows are being placed at jobs with titles that range from Quant to Data Scientist to Analyst to Statistician). Unfortunately, most hiring managers conflate the types of talent and temperament necessary for these roles.

While this isn't the only distinction among data scientists, it's one of the biggest when it comes to hiring. Here's the difference, and why it matters:

Analytics for machines: In this case, the ultimate decision maker and consumer of the analysis is a computer. Online ad or content targeting, algorithmic trading, product recommendations are just a few examples.

These data scientists are building very complex models ingesting enormous data sets and trying to extract subtle signals with machine learning and sophisticated algorithms. These digital models act on their own, choosing which ads to display, making recommendations to users, or automatically trading in the stock market, often in less than the blink of an eye.

Data scientists who produce analysis for computers need exceptionally strong mathematical, statistical, and computational fluency to build models that can quickly make good predictions. They usually operate with clear metrics (such as profits, clicks, purchases) and can piece together a myriad of technical tricks to build very sophisticated models that drive performance. When even small gains are aggregated across millions of users and trillions of events, their efforts can result in huge gains in revenue.

Analytics for humans: The ultimate decision maker and consumer of the analysis here is another human. Analyzing the effectiveness of products, understanding user growth and retention, producing reports for clients are just a few examples of the work these data scientists do.

They might be sifting through the same large data sets as their analytics-for-machines counterparts, but the results of their models and predictions are delivered to a human decision maker (often a non data-scientist) who has to make product or business decisions based on these recommendations.

Data scientists who produce results for people have to think about how to tell a story from the data. Because they have to explain their results to others — particularly those who are not as well versed in data science — they might deliberately choose simpler models over more accurate but overly complex ones. They have to be comfortable drawing higher-level conclusions — the “how” and “why.” These aren’t as easily observed in the data as the clear metrics enjoyed by their analytics-for-machines counterparts.

It’s important to get the right person for either job. We find that the typical profile of a data scientist who produces analytics for machines is someone with a natural science, mathematics, or engineering background (often at the PhD level) with the deep mathematical and computational background necessary to do the really high-powered work. Without the necessary technical skills, candidates will either fail at handling the large amounts of data or apply overly simplistic models that don’t capture the data’s full value.

However, these same people may not be suited to produce analytics for humans. Putting a team of MIT-trained physicists in a role where they are constrained to use “simple” models that management can understand will not be the best use of their talents, especially if they’re thirsting for a “deep” machine-learning challenge. On the other hand, social and medical scientists (again, often at the PhD level) are really well trained at understanding the “how” and “why” and often thrive on this kind of intellectual challenge.

Data scientists with hard science backgrounds have traditionally gotten a lot of attention in the press. In part, this is romanticizing the unknown — mysterious models that magically trade stocks or intuitively understand user preferences are sexier than the tedious work of thinking really hard about causality, sample bias, and the “how” or “why” of your data. But the latter could be what you really need from a data scientist. By asking this key question ahead of the hiring process, companies can go beyond the hype and find the right data scientists for their specific needs.

Michael Li is the founder and CEO of The Data

Incubator, a data science training and placement firm, which was acquired by Pragmatic Institute, where he is president. A data scientist, he has worked at Google, Foursquare, and Andreessen Horowitz. He is a regular contributor to VentureBeat, The Next Web, and *Harvard Business Review*. He earned a master's degree from Cambridge and a PhD from Princeton.