TECHNOLOGY

Big Data's High-Priests of Algorithms

'Data Scientists' Meld Statistics and Software for Find Lucrative High-Tech Jobs

Email   **Print**   **47 Comments**

By ELIZABETH DWOSKIN   CONNECT

Aug. 8, 2014 8:11 p.m. ET

Saba Zuberi, an astrophysicist working as a data scientist at TaskRabbit, said working for a consumer Internet firm can be surprisingly rewarding. *Ramin Rahimian for The Wall Street Journal*

For his Ph.D. in astrophysics, Chris Farrell spent five years mining data from a giant particle accelerator. Now, he spends his days analyzing ratings for Yelp Inc. YELP -1.86% 's online business-review site.

Mr. Farrell, 28 years old, is a data scientist, a job title that barely existed three years ago but since has become one of the hottest corners of the high-tech labor market. Retailers, banks, heavy-equipment makers and matchmakers all want specialists to extract and interpret the explosion of data from Internet clicks, machines and smartphones, setting off a scramble to find or train them.

"People call them unicorns" because the combination of skills required is so rare, said Jonathan Goldman, who ran LinkedIn Corp.'s LNKD -1.23% data-science team that in 2007 developed the
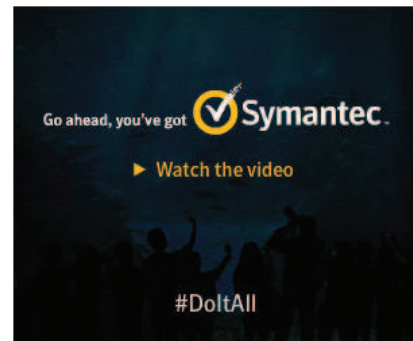
Popular Now   What's This?

ARTICLES

1 **Study: Half of Wildlife Lost Since 1970**

2 **Germs Thrive at Work, Too**

3 **Opinion: Obama Needs to Call Bush**

4 **The 2014 Baseball Playoffs Hateability Index**

5 **Planet Geek Meets Planet Jock**

"People You May Know" button, which five years later drove more than half of the invitations on the professional-networking platform.

Employers say the ideal candidate must have more than traditional market-research skills: the ability to find patterns in millions of pieces of data streaming in from different sources, to infer from those patterns how customers behave and to write statistical models that pinpoint behavioral triggers.

At e-commerce site operator Etsy Inc., for instance, a biostatistics Ph.D. who spent years mining medical records for early signs of breast cancer now writes statistical models to figure out the terms people use when they search Etsy for a new fashion they saw on the street.

At mobile-payments startup Square Inc., a Ph.D. in cognitive psychology who wrote statistical models to examine how people change their political beliefs now looks for behavioral patterns that would identify which merchants are more likely to have clients demand their money back.

Another 28-year-old at Yelp, with a Ph.D. in applied mathematics, turned his dissertation research on genome mapping into a product used by the company's advertising team. The same genome-mapping algorithm is now used to measure the effect on consumers when multiple small changes are made to online ads.

"Academia is slow and only a few people see your work," said Scott Clark, who designed the genome-mapping algorithm. "At Yelp, I can be pushing out experiments that affect hundreds of millions of people. When I make a small change to the Yelp website, I have a bigger impact."

Some such experiments have raised alarms. Facebook Inc. FB +0.05% was recently in the spotlight for an experiment in which its data-science team sought to manipulate people's emotions by altering the content of their news feeds.

Christian Rudder, president of IAC/InterActiveCorp.'s OkCupid dating site, recently disclosed in a blog post that the site manipulated its feeds by inflating the likelihood that any two people were a match, to encourage them to use the service more.

Mr. Goldman, who now heads a new data-science group at Intuit Inc., INTU -0.15% said employers go to great lengths to land top talent. They must be ready to extend an offer at a moment's notice, often within a day or two of interviewing a candidate, and be prepared to meet candidates at any hour of the day or night.

While a six-figure starting salary might be common for someone coming straight out of a doctoral program, data scientists with just two years' experience can earn between $200,000 and $300,000 a year, according to recruiters.

Anyone with "data science" in his or her job title on a LinkedIn page is going to get "100 recruiter emails a day," said Josh Sullivan, who leads a 500-person data-science group at the consulting firm Booz Allen Hamilton Holding Corp. BAH +0.34% To woo candidates, Mr. Sullivan goes for a personal touch: sending handwritten letters and flying across the country to meet potential employees' spouses. He also sends care packages filled with chocolate, as well as books on academic topics such as statistics and computer science that he knows the recruit is interested in.

VIDEO

1 **Aerial Drone Captures Scale of Hong Kong Protests**

2 **George Clooney, Amal Alamuddin Wed in Venice**

3 **Inside a Russian Billionaire's $300 Million Yacht**

4 **Five Things to Bring to the Occupy Central Protest**

5 **Online Community Helps Nerds, Geeks and Gamers Get Fit**

The scarcity is reflected in numbers. Job-listing sites SimplyHired.Com and LinkedIn currently list between 24,000 and 36,000 openings for positions that have data science in their titles. Data from a third site showed 6,000 companies were recruiting for such talent at the end of last year.

In 2012, the most recent year for which the federal government publishes such statistics, there were roughly 2,500 doctoral degrees awarded in statistics, biostatistics, particle physics and computer science—fields from which data scientists are typically recruited, according to the National Science Foundation. Over the past year, six universities, including the University of Virginia, Columbia University and Ohio State have launched or announced plans to launch certificate and master's programs in data science to fill the gap.

To get help, employers are increasingly looking to an elite program called Insight Data Science Fellows Program, which helps funnel doctoral candidates from fields such as astrophysics, neuroscience and math into the profession. The program, based near Stanford University and funded by tech companies, has a 100% placement rate.

Alums work in data-science teams at established Silicon Valley firms as well as startups such as Airbnb Inc., Palantir Technologies Inc. and Jawbone. This summer, the program expanded to New York City, where companies recruiting for data scientists include Viacom Inc. VIAB +0.16% 's MTV, Memorial Sloan Kettering Cancer Center, Capital One Financial Corp. COF -0.26% and the New York Times. NYT -2.43%

Some data scientists who five years ago would have gone to academia or become Wall Street quants said they felt the pull of the tech boom because funding for scientific research tightened up during the recession.

At the health-wearables company Jawbone, a data-science team headed by a former Insight fellow and computer science Ph.D. discovered that asking users to click a button called "Today I will" helped them meet their sleep goals. People who clicked the button, which had them promising to sleep a certain number of hours, went to bed an average of 23 minutes earlier than those who didn't.

Saba Zuberi, an astrophysicist working as a data scientist at TaskRabbit Inc., said working for a consumer Internet firm can be surprisingly rewarding.

At TaskRabbit, a startup that helps find hired hands for basic chores like packing boxes or housekeeping, users are shown a listing of potential "rabbits" who can do the tasks. To create the listings, Ms. Zuberi spent six months building a model that takes into account a worker's location, scheduling constraints, experience, ratings and payments rates, and attributes of the person making the request. The more factors that need to be weighted and matched, the more complex the model, she said.

Over time, the software learns which factors are more important to which customers and refines the listings. Ms. Zuberi said that while designing algorithms at TaskRabbit may not be as intellectually challenging as setting out to prove new theories of particle physics, the work felt more meaningful.

"At the end of the day, who you choose to show isn't just a listing," she said. "It's something that directly affects people's livelihoods."