Forbes / Tech

BETA

JAN 7, 2016 @ 01:47 AM 19,987

Big Data Uncovered: What Does A Data Scientist Really Do?

Bernard Marr, CONTRIBUTOR

FULL BIO

Opinions expressed by Forbes Contributors are their own.

Continued from page 1

Key qualities of a data scientist

In terms of personal qualities, a curiosity about data is essential, as well as communications skills, says Hanks. “People on my team spend a lot of time talking to customers to figure out what problems they need to solve, or talking to data vendors to find out what they can provide. So you become a middle man and communication is very important.”

Lots of different types of people go into data science, and Hanks explained to me that he was probably not a very typical example. However in my experience there is no such thing. The key capabilities Hanks mentioned cover a broad range of skills and people of different personality types and mind sets are attracted to the profession.

“I just really loved the interplay”, Hanks says, “From the beginning I was just totally fascinated. My first exposure to data science was probably in operations research, and I just loved the idea that you could take big data sets and use them to learn things, and improve things, and I found out that you really could use them to make a difference, I’ve found that fascinating for over 30 years now.”

Even after all that time in the business though, problems still come up which have him scratching his head, and these serve as a great example of the sort of challenges data scientists find themselves struggling with on a day to day basis.

“Just this morning I was working on something and one of the algorithms just wasn’t doing what it was supposed to do – basically it was showing us a link between a particular person

and a particular phone number which we just knew was incorrect. These problems can be very intermittent and very hard to diagnose.

"We have very specific algorithms that are supposed to do very specific things, and when they don't we just have to take them apart and find out why not, the problem is these days they are very complex and have a lot of working pieces! I can be completely mystified, like I am right now … but we will get there – we always do! That's really the sort of challenge we face day to day – systems which just don't behave the way they are supposed to according to our schematics."

In the time that Hanks has been working with data he has seen huge changes in the field, from working on structured databases on mainframes, to distributed Hadoop networks, to the cloud based, real time data processing world of today. So where does he see the future taking analytics and Big Data?

The Future of data science

Hanks sees a future of increased data streaming and real-time data processing, as opposed to huge batch processing of data. He believes that in this new world Hadoop MapReduce is less appropriate and in his work he is starting to use other systems like Scala and Akka.

One of the biggest challenges Hanks sees is the keeping up with the fast developments of new technologies and new algorithms. He believes that in order to be an effective data scientist you have to be holistic. He believes that it is relatively easy to become a specialist in MapReduce or a particular machine learning algorithm but the challenge is keeping up with the general speed of development in data science. "It's a field that is just stunningly big and complex, and has incredible breadth and depth", Hanks tells me, "You have to understand all of the pieces but the field is getting so vast – that's going to be the challenge facing data scientists going into the future."

Bernard Marr is a best-selling author & keynote speaker on business, technology and big data. His new book is Data Strategy. To read his future posts simply join his network here.