



HARI SARAVANABHAVAN

Hari is a senior business leader with 20+ years of work experience and has spent the last 13 years in the field of business analytics both as a customer and a service provider. Hari leads the business function within Data Sciences at Cognizant under the Digital Works umbrella. Hari is passionate about topics on embedded analytics, Digital analytics expansion and Data science evangelization.

There are a gazillion points of view on what data science is, its capabilities, and why being a data scientist is one of the sexiest jobs in America. There is also plenty of information on why we need data science and how it can do wonders for business. In a simple perspective, the digital world is transforming every single industry, and data science is at the core of that transformation.



One big question that gets asked frequently is this: What is the difference between data science and data analytics? There are many contradicting views (and rightly so) on which discipline does what. And history repeats itself. Remember the CRM wave that happened about a decade ago? It also meant different things to different people. Some thought of it as technology, some as interactions, and some as products.

Let's take a stab at clarifying. Data science has two key objectives:

1. Solve unresolved problems: This is more of a transformational or industry-level scenario, such as analytics in the Internet of Things (or Analytics of Things, as some call it).

2. Provide a different perspective to an existing problem: For example, how does telecom churn change in a digital world? The charter for data science has four major components

1. Data, data, data—any source, any way

While professionals may differ on this, in my opinion, having a fantastic handle on the data is a key differentiator and requirement for a good data scientist. What sources should we incorporate? Why should we incorporate those? The key here is to build a 360-degree view of the data. This is not to be confused with simple data management or the creation of tables or warehouses to store data. Instead, imagine a scenario where we have structured data, visual data, speech data, text data, and micro views on a particular topic. Building a 360-degree view of this is half the battle, and this is a key difference between a data scientist and a data analyst or miner. A data analyst or miner mostly focuses on what to do with the data and does not necessarily focus on what other data is important or look at alternate sources of data. All of this translates to a completely different set of skills. Functional knowledge, technological understanding, or domain expertise is what a data scientist brings to the discussion.



2. Make the data sources talk to each other

Sourcing the golden nuggets of data is one part, but the ability to link the various types of data together is a critical role of a data scientist. This is the second big difference in my opinion. A good data miner approaches this in a linear manner and may produce good results on each individual data set—for example, text mining on free text or sentiment analysis on social data. In the current digital world we need a nonlinear approach: one aggregated view on all the data sources combined, creating a robust sum of parts versus good output from every individual part.

3. Tools and techniques

Tools and techniques are needed to accommodate diversity in the data format, type, and size. With the 360-degree view of the data and the link in data sources, the traditional analytical tools and techniques will necessarily need a reboot to build the best output. Most people believe that focusing on this component (that is, machine learning or Python) is sufficient to become a good data scientist. Unfortunately, this approach takes away the “scientist” or exploratory thought process of the individual and results only in good execution. It is important for a sound data scientist to start with the data and follow through on execution.

4. Insights and interpretation

Another key trait that differentiates a data scientist from the rest is the ability to interpret the results and provide insights

to the executives that support real-world business impact and outcomes. While most assume these insights to be a given, the critical callout is the real-world business outcome versus expressing an internal analytics outcome. It is important to convert and relate the analytics throughput to understandable business measures. Without this relation, the output from data science will always be theoretical and not action oriented.

Becoming a master data scientist is a journey. All the components will play crucial roles in creating a data scientist that is more than science.

First published at
<http://digitally.cognizant.com/data-science-is-more-than-science>