

What is Data Science, Really?

Data science is hot stuff these days. You hear and read the term all the time. But what does it really mean? I heard one cynical professional say that data science is simply what consultants push after they fail to provide feasible use cases for block chain or big data or AI. Is that really all there is to it? Is it simply about burnishing resumes?

Let's face it though, the term 'data science' is indeed catchy. I think it's that word 'science'. We do believe in science, right? Well, I do – and I hope you do too. But a science of data? I mean where the data is literally about anything?! Really?! Clearly, we should take a closer look and nail down the underlying meaning (if any!). I give it a shot in this month's column.

Not too long ago I attended a conference on data analytics and machine learning. I listened to one innovative and exciting session after another. The term 'data science' was sprinkled generously throughout. But were they all talking about the same thing? Or was it simply a code word offering membership in an imagined community?

Indeed, 'data science' could simply be a term of convenience for a broad and enticing new marketing space. The industry loves that sort of thing. And *broad* indeed! Consider the following statement from Wikipedia (and pardon a nit about 'business' coming last):

"Data science ... incorporates skills from computer science, statistics, information science, mathematics, information visualization, data integration, graphic design, complex systems, communication and business."

I asked several times at the conference for a definition (which I admit, is a habit of mine – good or bad, depending on your point of view). I was

consistently disappointed. Perhaps 'data science' is a discipline that doesn't admit semantics(!?). That's a very interesting question, but I'll not digress.

The best response I got was something like, "It's more than statistics. More than business analytics. More than machine learning." To which was added, "You can't get an MBA to become a data scientist. Or get a degree in math. Or computer science."

Not an adequate definition at all! Not even a definition(!). Shared understanding does not arise from saying what something is not, or explaining how you can't become one. Shared understanding arises from expressing what something *is*. And no, you're not allowed to say, 'I know it when I see it'. I might look at the very same thing and *not* see what you see.

Wikipedia does give lots of good insights, but in my view falls short of a solid definition. It says:

"Data science ... uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured."

Some observations:

- It seems rather obvious that something called a 'science' would have to use 'scientific methods.' What's the real point there?
- The phrase "uses processes, algorithms and systems to extract knowledge and insights from data in various forms" could also be applied to 'data analytics' or 'statistical analysis.' What's the differentiation? It could also probably be applied to 'learning' (as in *machine learning*). What characteristic *specifically* makes data science a new, different thing of its own that can be clearly differentiated from similar and/or pre-existing things?

Wikipedia also says:

"The field encompasses preparing data for analysis, formulating data science problems, analyzing data, developing data-driven solutions, and presenting findings to inform high-level decisions in a broad range of application domains."

It's certainly true that we are seeing all kinds of new 'formulations' these days to develop 'data-driven solutions.' But that's sort of cherry-picking. What's the essence of the concept?

On social media, William Brooks suggested this definition:

data science: the application of the scientific method and experimental design to the statistical analysis of data

Much better! He went on to say:

- “This definition differentiates data science from most of the data analysis that goes on in business today.” (Yes, true, but correctly not part of the definition.)
- “Much in the same way that an Erlenmeyer flask might be used for scientific inquiry – or as a convenient vessel for a beer – machine learning uses tools of statistical analysis that may be used in data science or in other ways.”

That second point provides an excellent insight. A good ‘essence’ definition might therefore be:

data science: the application of the scientific method in using the tools of data analysis

That leaves one additional question. Can you really have a science of *data*? Has the world now become so digital that you can have a science of data about literally *anything*?

I suppose the answer depends on your definition of *science*. I hate to say it, but definitions provided by standard dictionaries support you either way. So, in the end (as always), the meaning is whatever the community says it is. I just wish the community would say it more clearly.

AUTHOR

Ron Ross

Ronald G. Ross is Co-Founder and Principal of Business Rule Solutions, LLC (www.BRSolutions.com). BRS provides workshops, consulting services, publications, and methodology supporting business analysis, business rules, business vocabulary, and rule management. His popular public seminars on business rules and business analysis, the first on business rules (starting in 1996) and the longest-running in the industry, are given through AttainingEdge (www.AttainingEdge.com). At BRS, Mr. Ross co-develops Proteus®, its landmark business analysis and business rules methodology, which features numerous innovative techniques including the popular RuleSpeak® (available free through www.BRCommunity.com). These are

the latest offerings in a 30-year career that has consistently featured creative, business-driven solutions. Mr. Ross also serves as Executive Editor of www.BRCommunity.com and its flagship on-line publication, *Business Rules Journal*. He is a regular columnist for the *Journal's Commentary* section which also features John Zachman, Chris Date, Terry Halpin, and Roger Burlton. *BRCommunity.com*, hosted and sponsored by BRS, is a vertical community for professionals working with business rules and related areas. Mr. Ross was formerly Editor of the *Data Base Newsletter* from 1977 to 1998. Mr. Ross is recognized internationally as the "father of business rules." He has served as Co-Chair of the annual Business Rules Forum Conference since 1997. He was a charter member of the Business Rules Group (BRG) in the 1980s, and an editor of the two landmark BRG papers, "The Business Motivation Model: *Business Governance in a Volatile World*" (2000, revised 2005) and the "Business Rules Manifesto" (2003). He is also active in OMG standards development for business rules and business models. Mr. Ross is the author of eight professional books. His newest are: *Business Rule Concepts* (2005), a second edition of his popular, easy-to-read 1998 handbook, and *Principles of the Business Rule Approach*, Addison-Wesley (2003), featuring the business rationale and opportunity for business rules. An earlier work, *The Business Rule Book* (1994, 1997), was the seminal work in the field. Mr. Ross received his M.S. in information science from Illinois Institute of Technology, and his B.A. from Rice University.