

What is ‘data science’ / data analytics? - Yet another opinion

fortune Walla

June 20, 2015

Created on Friday, February 21st, 2014 at 6:52 am

In the last twenty years of the Information Age, computer networks & the Internet have allowed us to gather & store information from a wide variety of devices (industrial/medical equipment, photos/audio/video sensors, ATMs, credit/debit cards, mobiles & computers, etc..) & platforms (retailing, telecom, banking & insurance , pharmaceutical, security, social media, etc..).

The information being collected is increasing as more people & businesses use computers & the Internet as a primary means for conducting business, social & monetary transactions. Technologies such as 3G/4G data networks, mobile computing devices & cloud based computing systems have accelerated this trend.

“The value of data is no longer in how much of it you have. In the new regime, the value is in how quickly and how effectively can the data be reduced, explored, manipulated and managed.”

Usama Fayyad – President & CEO of digiMine, Inc. [1]

Questions to think about:

- Can the increasing amount of information be collected, stored & managed in a consistent manner for easy access to others?
- Can information from different sources be integrated & linked to create meaningful relationships between them?
- Can information systems be built to easily get answers to characteristics of stored data?
- Can we detect underlying patterns in the data & “mine” the data to reveal patterns of behavior that will be insightful & profitable in terms of business/research before it loses its value?
- Can analytic processes/models be built to allow prediction of future outcomes/behavior from existing data more quickly than the competition?
- Can it be ensured that business/research returns will be greater than the investment cost of the data collection & analytics system?

It takes brilliance to ask the right questions, at the right time in history. The value of a Big Data resource is that a good analyst can start to see connections between different types of data, and this may prompt the analyst to determine whether there is a way to describe these connections in terms of general relationships among the data objects [2]

It is important to realize that most of these problems have been discussed [1] & studied in research journals & other industry publications for the last 30-40 years under various labels such as “Business Intelligence”, “Knowledge Discovery”, “Data Mining”, “Decision Science”, “Statistical Learning”, “Predictive Modeling”, “Machine Learning”, “Business Forecasting”, etc... Essentially, it is the coming together of data analysis techniques, large scale computing & domain knowledge.

Now in the present time, all these are under the labels of “Big Data” [2] “Data Science” & “Data Analytics”. The big change now is in the commoditization of technologies where all these techniques can be applied in a cost effective manner & almost in real-time.

While it is not realistically possible for a single person to perform all the above tasks, a new practitioner has emerged who has the relevant knowledge in statistical & data mining techniques, computing & programming techniques as well as domain knowledge of the business/research problem.

The main skill required (besides the usual technical knowledge) is curiosity about what patterns exist or can be “mined” from stored data sources as well what can be predicted from it, the ability to experiment with new methods to get new insights & explanations just like a scientist would. Perhaps that is why they call people in this field “data scientists”!

My own equation for this “emerging” field would be:

Data Science/Data Analytics =
Domain knowledge of the business/research problem +
Mathematical formulation of business/research problem into a statistical model +
Programming the statistical model into software code +
Business/Research analysis of the statistical output

References

- [1] HAMPARSUM BOZDOGAN (ed.) *Statistical data mining and knowledge discovery* (2003) Chapman & Hall/CRC
- [2] JULES J BERMAN *Principles of big data: preparing, sharing, and analyzing complex information* (2013) Morgan Kaufmann