

BUSTING data science myths

SAIM AHMED SHIFAT

It is a universal truth that all modern businesses today are more or less packed with a lot of data. 1 Exabyte (1^{018}) of data is being created on the internet daily. With the emergence of the Internet of Things (IoT), more organisations today are opening up their doors to big data and unlocking its power thereby increasing the value of a data scientist who knows how to generate actionable insights out of gigabytes of data. Besides organisations, some countries also are taking initiatives to incorporate data science analysis to save costs. According to a report by McKinsey & Company, big data initiatives in the US healthcare system could account for USD 300 billion to USD 450 billion in reduced healthcare spending or 12 to 17% of the USD 2.6 trillion baselines in US healthcare costs.

Data science continues to evolve as one of the most promising and in-demand career paths for professionals. Advancing past the traditional skills as to analysing data, data mining, and programming skills, successful data professionals today are mastering the full spectrum of the data science life cycle.

With increasing popularity, Big Data and Analytics are two of the most common terms today that are mentioned in any product introduction by a company, career talk or skill development sessions. However, very few people know what Big Data and Analytics actually are and what the differences are between the two. Moreover, the plethora of contents and blogs on the internet have led to

inconsistent definitions and some myths regarding data science.

Data Science is only for people with an engineering background

K. M. Saqiful Alam, an expert in analytics and machine learning, is currently pursuing his PhD to explore links between big data analytics, firm-level strategy, and behavioural biases. He completed his undergraduate studies in business. When asked about the myth, he said that anyone could be good at data science irrespective of their academic background if he/she follows the right resources under proper guidance. According to him, all the data around us end up telling a story. Every bit of data has something to do with the story of everything around us. The urge to understand and tell these stories is very important and one does not necessarily have to be from an engineering background to do that.

Data Scientists are not coders

Different types of media today promote self-claimed data scientists who do not know how to code. Although Excel, SPSS and Tableau can help one by creating visualizations to impress the viewers, these tools are never enough to reach the depth of analytics and prediction based on collected data. As a start, people usually start by learning to code with R or Python. Websites like DataCamp are dedicated to teaching people about coding relevant to analytics and machine learning.

Data Science requires a deep understanding of Statistics and Statistical Methods

Data Science requires an understanding of statistics. However, businesses can take advantage of Data Science without having a dedicated statistician on staff. One does not have to take a formal course or pursue a graduate degree for this. There are plenty of e-books and other resources on the internet that helps understand the basics of relevant statistics. By dint of this learning, one can build models that are meaningful to his/her organisation.

Analytics, Data Science and Machine Learning are the same things

Today all these buzzwords are being used everywhere on the internet interchangeably without actual explanations of which is what. Some have incorrect definitions of the terms while others think these are different ways of calling the very same thing. Saqiful Alam explained this conundrum by explaining the role of each of the components in the whole process. It starts with Analytics, which can be of three types, viz. Descriptive, Prescriptive (use of statistics) and Predictive (use of statistical models like regression). Predictive analysis is the area where Machine Learning (ML) is used the most. Companies and developers put increasing efforts behind making these processes as automated as possible. And this is where AI (Artificial Intelligence) comes in. In all of this, according to Saqif, "Data Science helps in the derivation stage, not the implementation".

When asked, Saqif mentioned the following steps to be helpful for anyone who wants to be an Analyst:

- **Understanding that it is all about a story in the first place:** The skill of identifying the right story to tell with the data is very important. This, in turn, will guide the learners towards the data they want to collect.

- **Understanding the power of stats:** Statistic is important for data analysis and the following resources can help the learners with that

- a. Statistics with R Specialization: Online set of courses by Duke University in Coursera.

- b. OpenIntro Statistics: A book on statistics offered as a supplement of the mentioned course.

- c. Youtube Channels: Khan Academy and StatQuest with Josh Starmer.

- d. Naked Statistics: A book that talks about connecting Statistics with the story analysts want to tell.

- **Taking small baby steps towards coding:** R or Python both the languages are very much in use and learning one can speed up the process of learning the others. A very good source for learning coding relevant to Data Science in small steps is DataCamp.

- **Enjoying learning to be a Data Scientist:** Aiming to be a Data Scientist just because everyone is talking about it or because it sounds cool/sophisticated will not take anyone far enough. To tell a unique story and stand apart from the crowd, enjoying Data Science and being curious about it is very important.



**4-LANE DHAKA-BHANGA
EXPRESS HIGHWAY
BUILT WITH BS RM XTREME**

COUNTRY'S NO. 1 STEEL EXPERT

