

'50 Years of Data Science, David Donoho'

An assignment by MM73053

Summary -

50 years back, Prof. Jhon turkey showed the gap where data analysis is an unexplored part of science. Similar, professors -John Chambers, Bill Cleveland, and Leo Breiman highlighted the need for statistics to expand its boundaries beyond theoretical statistics and emphasized the importance of data preparation and preparation for data modeling and also on the prediction of data rather than just inferencing the same. Hence, the name 'Data Science was coined' by Cleveland. To current contemporary definition Data science supersedes with some fields of stats and machine learning, Data analysis while having big data as part this science.

The data science and Statistics conundrum -

Various organizations and stats scholars began reacting to above scenario saying the definition of statistics encompasses to that of data science, as some say stats don't just inference a small amount of data, they do it irrespective of size. And they used to do much more than just inferencing the data by their data modeling techniques. However, there are other sets of organizations and scholars who exclaim stats are just a subset of data science, which I strongly agree about. This notion is explained in terms of Big Data is different from stats as the technology leverages exclusively on handling and maintaining a plethora of data. This differs from stats, wherein they've been used only to derive the inferences from the data that's been out on the problem table.

Gaining Skill level for a data scientist goes post master's degree, learning and honing on skill development for years before they can add value to employer's organization. A good data scientist should not be afraid to get their hands dirty and try to become one with the data by exploring it.