50 years of Data Science

The author in this interesting article is helping us understand the field of Data Science from its inception to the future and how it differs with the traditional statistics.

Firstly, he describes about the present state of Data Science which is gaining importance in almost all fields clearly seen by number of courses coming up and the grants allocated to it in many of the prestigious institutions. The University of Michigan’s Data Science initiative which in a press release said it had become the fourth approach to scientific discovery along with experimentation, modeling, and computation.

Next, he brings about the comparison between Data Science and statistics along with prominent memes contributing to this field. He also cites articles where it shows that Statistics is a subset of data science and contrary to this a few articles on Data Science without the use statistics. Evidently there are a number of interpretations of Data Science and its relation to Statistics. Big data which deals with large datasets is being part of Data Science which uses several statistical sampling whereas the earlier statistics was defined for the use of large datasets but was used among smaller datasets for the ease of understanding. Furthermore, the skillsets of dealing with large scale organizational datasets is in demand not the conventional problem-solving ability to find insights from the data. Moving forward with the jobs scenario there is a dearth of people having skills in both databases and statistics. Then he briefly explains on what a good data scientist and a successful data scientist mean.

Lastly, he says as the world is moving forward with technological advancements is going to generate vast amount of data for which a proper framework should be worked out to make effective use of the data.