Louise Bassog

Dr. Bemley

CTEC 298

10/25/2022

**What Is Data Science?**

A data scientist is a person who can extract meaning from and analyze data. This involves both the tools and methods from the fields of statistics and machine learning, in addition to the fact that the data scientist must also be human. Since data is never in its purest form, she must spend a significant amount of time gathering, cleaning, and manipulating it. This method calls for tenacity, knowledge of statistics, and expertise in software engineering. These are the same talents that are required for comprehending any biases present in the data as well as for troubleshooting the logging output from code.

Exploratory data analysis, which combines visualization and data sense, is a vital element of the process once she has gotten the data into shape and organized it. She will uncover patterns, construct models, and develop algorithms; some of these will be developed with the goal of comprehending how the product is used and the state of the product, while others will act as prototypes for features that will eventually be baked back into the product. She may design experiments, and she plays an essential role in the process of making decisions based on evidence. She will communicate with team members, engineers, and leadership in a straightforward manner and with data visualizations so that even if her colleagues are not immersed in the data themselves, they will comprehend the significance of what is being communicated.

Because no one is exceptional in every area, the most effective way for a data science team to function is for its members to possess a diverse set of skills, including those in computer science, mathematics, statistics, machine learning, domain expertise, communication and presentation skills, and data visualization. This leads us to speculate about whether it would be more beneficial to define a "data science team." Hacking driven by Red Bull and statistics powered by espresso — that's what data science, as it's now conducted, is all about. But data science is not simply hacking—because when hackers are done, data scientists have something to show for it. Few of them are concerned with non-Euclidean distance measurements while they are debugging their one-liners in Bash and their Pig scripts. And statistics is not the same thing as data science since by the time statisticians have finished conceptualizing the ideal model, very few of them would be able to read a tab-delimited file into R even if their jobs relied on it.

Data science may be thought of as the field's equivalent of civil engineering. Its adherents are equipped with both a hands-on mastery of various instruments and materials and a theoretical comprehension of the range of possibilities available to them.