PART II

Programming Foundations for Data Science

CHAPTER 8

What Is Data Science?

Data science encompasses the tools and techniques for extracting information from data. Data science techniques draw extensively from the field of mathematics, statistics, and computation. However, data science is now encapsulated into software packages and libraries, thus making them easily accessible and consumable by the software development and engineering communities. This is a major factor to the rise of intelligence capabilities now integrated as a major staple in software products across all sorts of domains.

This chapter will discuss broadly on the opportunities for data science and big data analytics integration as part of the transformation portfolio of businesses and institutions and give an overview on the data science process as a reusable template for fulfilling data science projects.

The Challenge of Big Data

Due to the expansion of data at the turn of the twenty-first century epitomized by the so-called 3Vs of big data, which are volume, velocity, and variety. Volume refers to the increasing size of data, velocity the speed at which data is acquired, and variety the diverse types of data that are available. For others, this becomes 5Vs with the inclusion of value and veracity to mean the usefulness of data and the truthfulness of data, respectively. We have observed data volume blowout from the megabyte (MB) to the terabyte (TB) scale and now exploding past the petabyte (PB). We have to find new and improved means of storing and processing this ever-increasing dataset. Initially, this challenge of storage and data processing was addressed by the Hadoop ecosystem and other supporting frameworks, but even these have become expensive to manage and scale, and this is why there is a pivot to cloud-managed, elastic, secure, and high-availability data storage and processing capabilities.

E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*,
https://doi.org/10.1007/978-1-4842-4470-8_8