

Career in Data Science: What you need to do in college for a strong future?

Dr. Sanjay Goel

Data plays a very important role for understanding our world, physical, biological, social, economic as well as ourselves. Availability of data usually reduces uncertainty and ambiguity in our understanding. Our dependence on data for clearer understanding and decision making increased with the scientific as well as industrial revolutions. With increased trust on data, seven hundred years ago, the world's data growth curve started becoming more vertical and five hundred years ago modern statistics, mostly related to population data, started to inform the policy and even reform the society. Nearly seventy years ago, computers started to transform our ability to encode, store, analyse and later even share data. Around thirty years ago, internet started to contribute to create a flood of data. Broadband networks, mobile devices, streaming media, and IoT on one hand, and trends like urbanisation, e-governance, e-commerce, online learning, e-health, open government data, etc. on the other, are now rapidly transforming this flood into an overwhelming deluge.

In 2012, we started entering into a zettabyte age when world's total digital data exceeded one zettabyte, i.e., a trillion gigabyte. As per some estimates, the global data volume will touch 180 zettabytes by 2025, 572 zetta byte by 2030, and it may even go up to 5,00,000 zetta byte by 2050, with 90% data being real-time. Data has become more important than ever across industries today, leading to a growing demand for data science professionals who are playing a pivotal role in helping businesses and organisations generate actionable insights. Converting this voluminous data will require a great deal of effort by data scientists, however, technologies like Machine Learning and predictive analytics are making the process faster as well as more efficient. While dealing with such huge quantities of data will be a challenge, it will create a lot of opportunities and jobs as well.

What is data science?

In simple terms, data science encompasses preparing data for analysis, including cleansing, aggregating, and manipulating the data to perform advanced data analysis. It is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured as well as unstructured data. According to a 2021 ACM report on Data Science, it is the field that brings together domain data, computer science, and the statistical tools for interrogating the data and extracting useful information. Domains create and provide the structured as well as unstructured data; statistics and mathematics provide tools for analysis, modeling, and inference; and computer science contributes algorithms and technologies for data access, management, protection, as well as effective processing in modern computer architectures.

What do data scientists do?

Data scientists have become necessary assets and are present in almost all organisations nowadays. These data-driven professionals identify problems or objectives that can be addressed through data analysis by identifying

relationships and trends that help in developing solutions to business, engineering, governance, or scientific problems using mathematical theories and techniques and recommend data-driven solutions to key stakeholders. This filtered information can be used to answer questions and drive strategy in an organisation or business. In this process, they clean and process large sets of data, apply feature selection algorithms, compare, test, validate, and reformulate models, create visualisations, and write new programs or applications to conduct analyses. They may even have to design instruments to collect data.

What are the career prospects in data science, globally and in India?

According to Michael Page's July 2021 report, the rise of data science needs will create 11.5 million jobs opening by 2026 globally. According to Analytics India Magazine, in 2022, India contributed 11.6% of the total open job globally in data science and analytics. With a large pool of STEM graduates, there is huge opportunity for India to emerge as a world leader in this field. Analytics India Magazine's June 2022 report estimates that overall Analytics industry will exceed US\$ 200 billion by 2027 and Indian companies are focusing on increasing the share of fresh recruits to increase their share in this growth.

What should students learn for a bright career in data science

In order to serve in the areas of data analytics, data science, or business analytics in business, consultancy, industry, government, healthcare, education, research, etc., students should basically develop the abilities to:

1. identify, extract, and pull together available and pertinent heterogeneous data and use appropriate computational principles, platforms and technologies to discover new relations and deliver insights into research problem and/or organisational processes.
2. conceive, design, implement, and manage data analytics, data management and information systems, services, and processes by using principles of computer science, data management, machine learning, computational statistics, software engineering, and state of the art platforms, components and tools.

A good UG education in computer science, engineering, statistics, or mathematics prepares the students with the fundamental knowledge and skills required to develop above abilities through further education or training. Many universities are now offering MTech, MSc, MBA or PG Diploma programs in Data Science, Data Analytics, or Business Analytics. A good technology focused master's program in data science will normally include courses like computational statistics, data mining, machine learning, deep learning, big data, cloud computing, web analytics, large graph analytics, natural language processing, computer vision, software engineering, etc. Depending upon their strengths in a few application domains, some programs may also include a few domains specific analytics courses like financial analytics, ecommerce analytics, manufacturing analytics, sales and marketing analytics, transport analytics, tourism analytics, HR analytics, health analytics, agriculture analytics, etc. Some universities are also offering UG programs in these areas. Many online programs are also available but online programs usually work only for the highly self-motivated students. However, online courses can surely be a great supplement to regular university programs. Some universities have already partnered with online platforms like Coursera or edX to strengthen their students learning.

Students should focus on building their knowledge and skills in statistics, algebra, calculus, programming, machine learning, big data, cloud computing,

and software engineering. They should learn computing platforms, technologies and tools like Linux, Python, R, Java, SQL, Tableau, TensorFlow, Apache Spark, Hadoop, Docker, etc. They must sharpen their critical thinking to ask interesting questions and understand issues and situations from multiple perspectives. They need to develop their interpersonal skills to communicate with others about complex processes and possibilities. They should also learn to accommodate themselves well in team work.

Data science is crucial at a time when India aims to become a global hub for manufacturing, Artificial Intelligence (AI) and related technologies. A large number data science jobs are set to be in demand in India as well as overseas in the coming years and students interested in this field should acquire adequate skills while they are in college to build a bright career for themselves. Also, research on expertise has shown that it normally takes around ten years of rigorous engagement and deliberate practice in any field to become an expert. So students must have patience, remain motivated, and enthusiastically engage with the discipline to develop their knowledge and skills. They must also supplement their academic studies with participation in professional networks, conferences, competitions, boot camps and engagements in projects and internships.

Disclaimer

Views expressed above are the author's own.

END OF ARTICLE