



50 YEARS OF DATA SCIENCE

Shah Vishruti¹, Hiren Mer²

¹ B. Tech Student, Department of CSE, Indus Institute of Technology, Indus University, Ahmedabad, Gujarat, India

² Assistant Professor, Department of CSE/IT, Indus Institute of Technology, Indus University, Ahmedabad, Gujarat, India

Abstract

Data science is the advent of scientific studies of data analysis on a scientific scale, not only scaling up. This review will provide an in-depth analysis of the development of data science, including its survey-based expansions and growing demand, as well as its practical significance. This page presents a quick overview of data science, including its application and potential future research areas.

Keywords:- Data Science, Data Analysis, Methodologies

Introduction

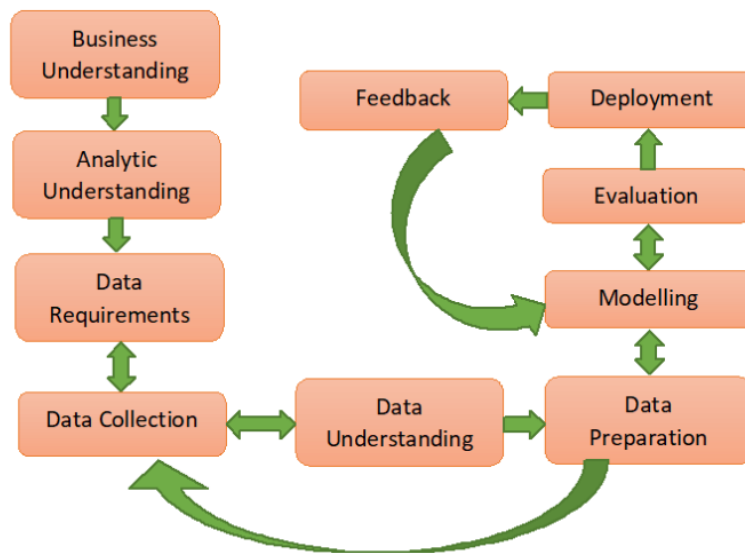
C.F. Jeff Wu created the term "data science" to replace the term statistics, while John Turkey defined the discipline he dubbed "data analysis" in 1962. This area and contemporary data science are comparable. Simply put, "the concept of 'Data Science' is introduced by the coupling of scientific discovery and practice, which involves the collection, management, processing, analysis and visualization, and interpretation of vast amounts of heterogeneous data associated with a diverse array of scientific, translational, and interdisciplinary actions." In data science, modeling, statistics, engineering, economics, and analytics are merged with the power of computer science. A data scientist will most likely look into and examine data from numerous sources. Data science builds upon the first three methods of scientific discovery, which were modeling, computation, and experimentation.

History And Background:

A few academic institutions have developed creative programs in response to the growth of data science. Many entering students at the University of California, Berkeley are now provided with connector courses covering cultural data, social data, demography, small cities, ethics, and social networks in addition to the Data 8 introduction source. Although data science is frequently referred to as a young field, mathematicians have been working in the field for decades. Donoho outlines the emergence of data science as a field and makes predictions about how computational and inferential techniques will be combined with statistics and machine learning in the future.

Data Science Methodology:

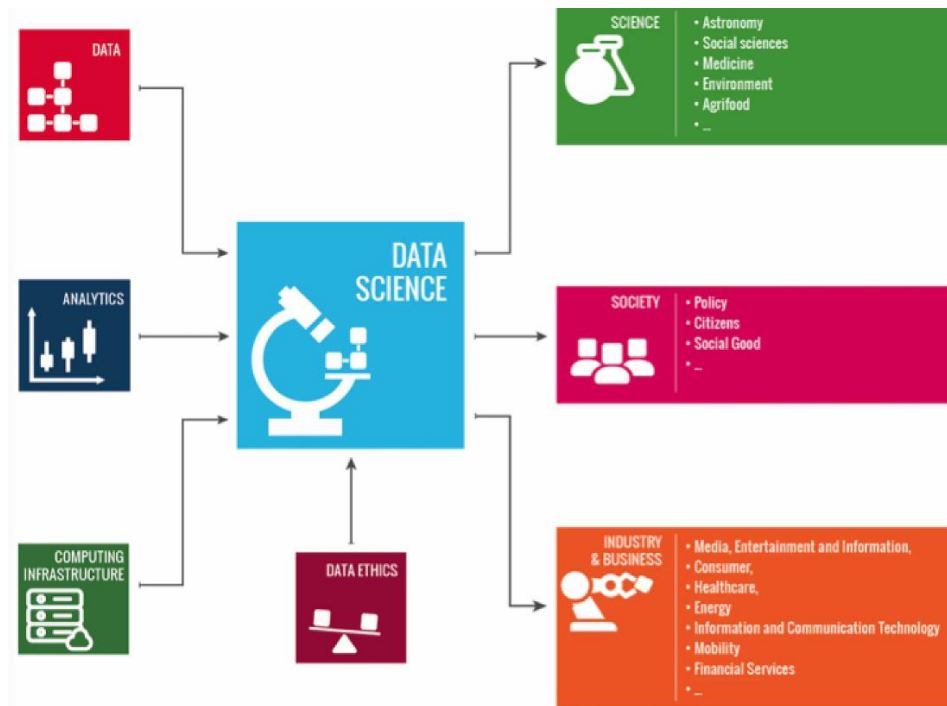
The Data Science Methodology outlines the process for solving a particular issue. This is a cyclical process that is subject to critic behavior, which directs data scientists and business analysts to take appropriate action.



- 1) Business Understanding- Before solving any problem in business domain, it needs to be understood properly.
- 2) Analytic Understanding- Based on the above business understanding, one should decide the analytical approach to follow.
- 3) Data Requirements- The above chosen analytical method indicates the necessary data content, formats and sources to be gathered.
- 4) Data Collection- Data can be collected in any random format. So according to the approach chosen and the output to be obtained, the data collected should be validated.
- 5) Data Understanding- Descriptive statistics calculates the measures applied over data to access the content and quality of matter.
- 6) Data Preparation- It includes transformation, normalization, etc.
- 7) Modelling- Modelling decides whether the data prepared for processing is appropriate or requires more finishing and seasoning.
- 8) Evaluation- Model evaluation is done during model development. It checks the quality of model to be assessed and also if it meets the business requirements.
- 9) Deployment- As the model is effectively evaluated, it is made ready for deployment in the business market.
- 10) Feedback- Feedback is necessary purpose which helps in refining the model and accessing its performance and impact.

Research Through Innovation

Data Science for Society, Science, Industry And Business:



The following scientific, technological, and socioeconomic variables interact in the ecosystem that is data science:

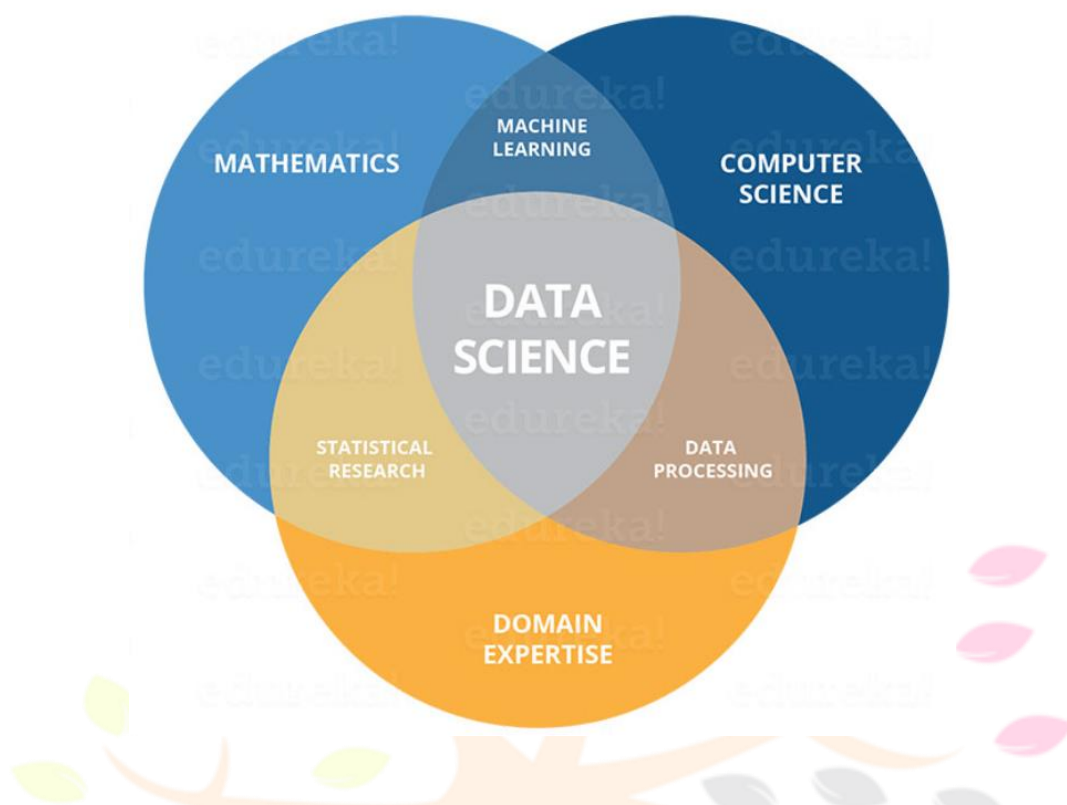
- i) Information Data accessibility and availability;
- ii) Infrastructures for computing and analytics accessibility to open-source analytics and high performance analytical processing;
- iii) Competencies Availability of highly qualified engineers and data scientists;
- iv) Legal and ethical considerations Regulations pertaining to data ownership and use, privacy and data protection, security, liability, cybercrime, and intellectual property rights are available;
- v) Usage Applications for business and the market;
- vi) Social aspects Concentrate on the main worldwide societal issues.

Some common examples are:

- i) The development of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has revolutionized image recognition and natural language processing.
- ii) Bias in machine learning models has raised concerns, leading to efforts to develop fair and ethical AI systems.

Importance

Data science has advanced significantly in recent years, making it a crucial component in comprehending how different industries operate..

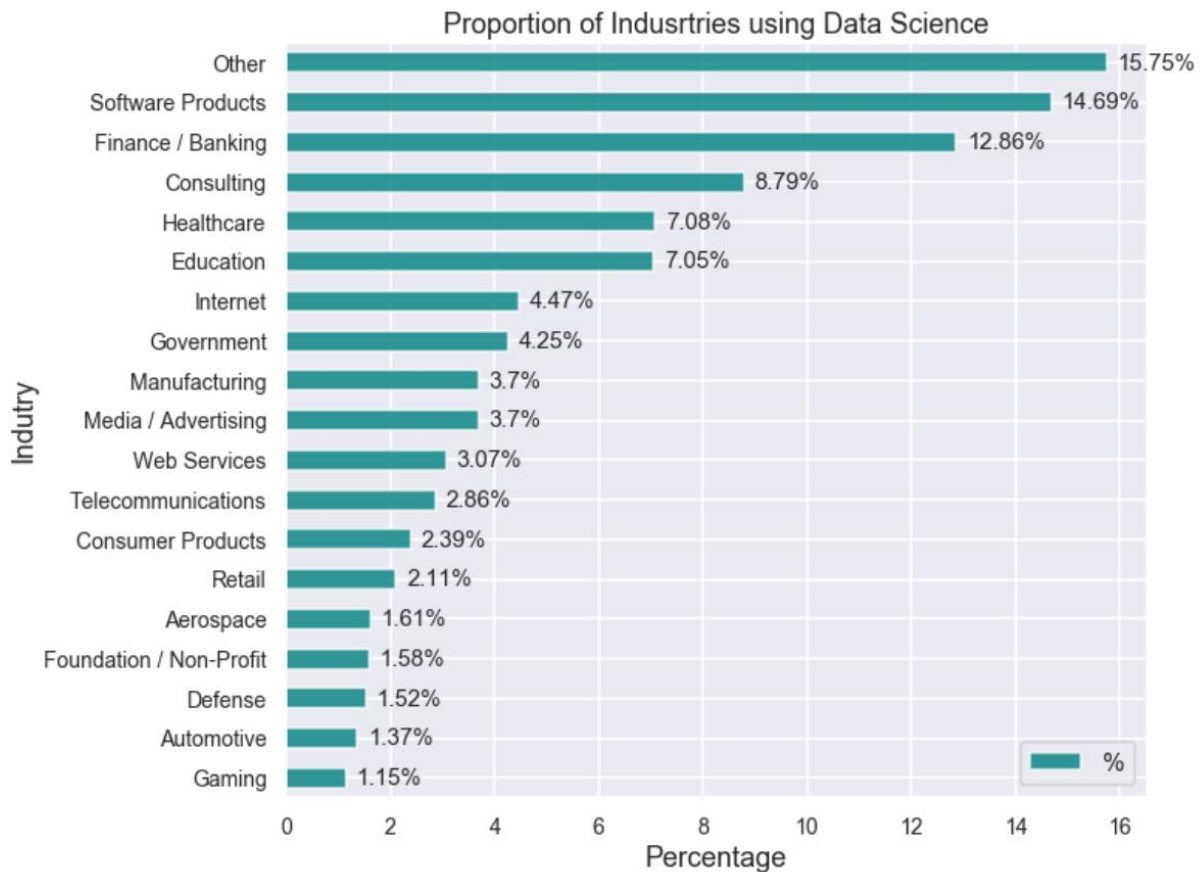


The expanding scope of data has brought the multidisciplinary field of data science into focus and increased its importance. It makes use of scientific procedures, methods, algorithms, and structures to extract knowledge and insight from vast amounts of data. This idea combines concepts, data analysis, machine learning, and related strategies with data to analyze and comprehend real-world phenomena. Data mining, statistics, predictive analysis, and other aspects of data processing are all expanded upon by data science. The vast field of data science employs numerous methods and ideas from other data fields, including computer science, statistics, information science, and mathematics..

Next 50 years Of Data Science:

Technology has the great quality of demanding. It requires progress. Starting in 2023, we will produce more data per minute in the next 30 to 50 years than we do in a day at the moment. Companies will require sophisticated computers to perform quantum computing on such data, as standard computers will not be sufficient to handle it. Companies will need to process enormous amounts of data even for simple analysis tasks because it will be difficult to clean and mine the data to make it structured.

Here's a tiny sample of the industries that use data science:



Thus, based on the information above, we can conclude that data science will have an enormously exciting future due to the introduction of new techniques.

Summary:

Our perspective on data science is comprehensive and has implications for research, practice, and education. It also highlights the necessity, methods, and significance of data science in our daily lives. It is essential for boosting desired behaviors within a system. In the upcoming decades, as scientific data and data about science itself become more widely available, the scope of this science will greatly expand.

References:

- 1) <https://www.tandfonline.com/doi/epdf/10.1080/10618600.2017.1384734?src=getftrhttps://www.tandfonline.com/doi/full/10.1080/10691898.2020.1851159>
- 2) https://link.springer.com/article/10.1007/s41060-020-00240-2?utm_source=getftr&utm_medium=getftr&utm_campaign=getftr_pilot
- 3) <https://www.pnas.org/doi/pdf/10.1073/pnas.1702076114>
- 4) <https://com-mendeley-prod-publicsharing-pdfstore.s3.eu-west-1.amazonaws.com/>
- 5) <https://www.educba.com/data-science-and-its-growing-importance/>
- 6) <https://www.geeksforgeeks.org/data-science-methodology-and-approach/>
- 7) <https://qubeshub.org/publications/912/1>
- 8) <https://www.tandfonline.com/doi/full/10.1080/10618600.2017.1384734>
- 9) <https://www.dataversity.net/50-years-of-data-science-what-it-is-and-what-it-can-be/>
- 10) https://www.mendeley.com/search/?query=50+years+of+data+science&dgcid=md_homepage
- 11) <https://www.mendeley.com/catalogue/225f1ef5-7669-39cd-89f3-000ce0ba479d/>
- 12) <https://www.mendeley.com/catalogue/eecd7c09-a9be-397c-9e95-4aff4a104b52/>