

Data Science

OPEN ACCESS

Manuscript ID:
ASH-2022-10014969

Volume: 10

Issue: 1

Month: July

Year: 2022

P-ISSN: 2321-788X

E-ISSN: 2582-0397

Received: 15.05.2022

Accepted: 10.06.2022

Published: 01.07.2022

Citation:

Sankar, K. "Data Science." *Shanlax International Journal of Arts, Science and Humanities*, vol. 10, no. 1, 2022, pp. 86–91.

DOI:

<https://doi.org/10.34293/sijash.v10i1.4969>



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

K. Sankar

Assistant Professor of Computer Science

Lal Bahadur Shastri Govt. First Grade College, R T Nagar , Bangalore, India

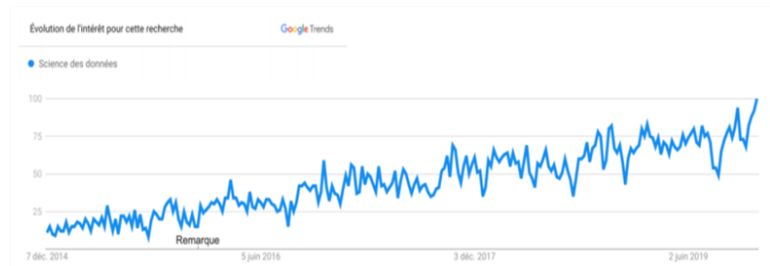
Abstract

Data science has become the most demanding task of the 21st century. All companies are looking for candidates with knowledge of data science. This topic provides an overview of data science. Includes data science duties, data science tools, data science components, applications and more.

Keywords: Data, Science, Tools, Technology, Applications

Introduction

Data science crosses paths with both big data and artificial intelligence when it comes to analyzing and processing data called datasets. Google Trends is a great way to show that data science as an intended topic has increased over the years since 2004.

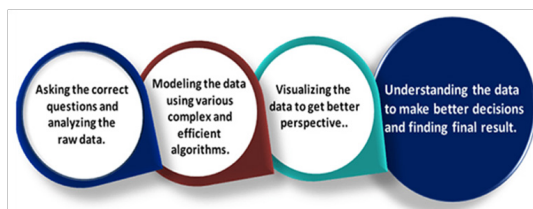


What is Data Science?

Data science is a detailed study of vast amounts of data, extracting meaningful insights from raw, structured, and unstructured data processed using scientific methods, various technologies, and algorithms. To do this is an interdisciplinary field of manipulating data using tools and techniques, where you can find new meanings. Data science uses the most powerful hardware, programming systems, and efficient algorithms to solve data-related problems. It's the future of artificial intelligence.

In short, data science is like this:

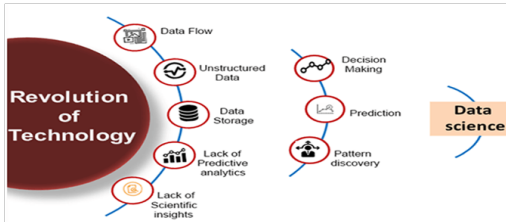
- Ask the right questions and analyze the raw data.
- Modeling data using a variety of complex and efficient algorithms.
- Visualize your data to get a better perspective.
- Understand your data, make better decisions, and find revenue.



Example

Suppose you want to drive from station A to station B. Now we need to make some decisions. B. Which route is the best route to get to the place faster and which route has less traffic and is cost-effective.

All of these determinants act as input data, and you get the right response from these decisions. Therefore, this data analysis is called data analysis, which is part of data science. The Need for Data Science:



Years ago, data was scarce and mostly available in a structured format that could be easily stored in an Excel spreadsheet and processed by BI tools. But in today's world, the amount of data is increasing. H. Approximately 2.5 quintals of data bytes are generated each day, leading to an explosive growth of data. By 2020, it is estimated that one person on earth will generate 1.7 MB of data per second. Every company needs data to make its business work, grow, and improve. Processing such huge amounts of data today is a daunting task for all businesses. Processing, processing, and analyzing this required complex, powerful, and efficient algorithms and technologies. This technology has emerged as data science. The main reasons for adopting data science technology are:

With the help of data science technology, vast amounts of raw and unstructured data can be transformed into meaningful insights. Data science technology is opting by various companies, whether it is a big brand or a startup. Google, Amazon, Netflix, etc, which handle the huge amount of data, are using data science algorithms for better customer experience.

Data science is committed to automating transportation, including the development of self-driving cars, the future of transportation. Data science can help you make different predictions, such as different votes, elections, ticket confirmations, and more.

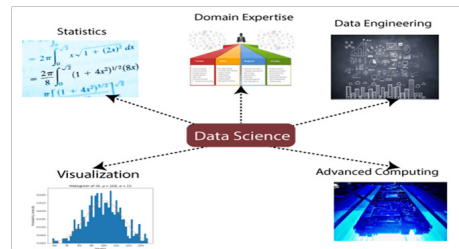
Data Science Job Types

Learning data science gives you the opportunity to find a variety of exciting work roles in this area. The main duties are as follows.

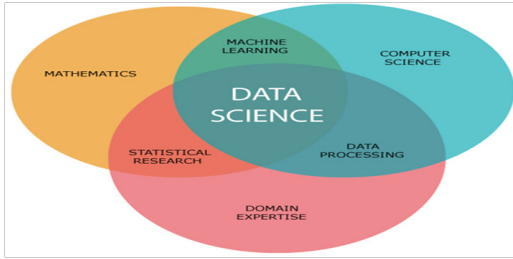
1. Data Scientist
2. Data analyst
3. Machine learning expert
4. Data engineer
5. Data architect
6. Data controller
7. Business analyst
8. Business Intelligence Manager

Data Science Components

The main components of data science are:



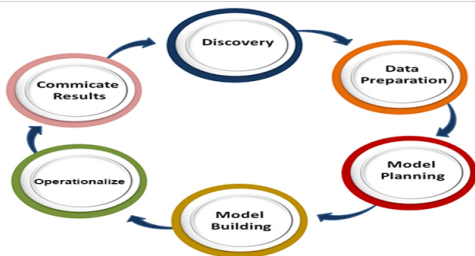
1. **Statistics:** Statistics are one of the most important elements of data science. Statistics are a way to collect and analyze large amounts of numerical data and derive meaningful insights from them.
2. **Domain expertise:** In data science, domain expertise connects data science. Domain expertise means expertise or skills in a particular area. There are several areas in data science that require domain experts.
3. **Data Engineering:** Data engineering is part of data science, including data collection, storage, acquisition and transformation. Data engineering also includes metadata about data (data about data).
4. **Visualization:** Data visualization is the display of data in a visual context so that people can easily understand the meaning of the data. Data visualization gives you easy access to the vast amount of data in your visuals.
5. **Advanced Computing:** An important element of data science is advanced computing. Advanced computing involves designing, creating, debugging, and maintaining the source code of computer programs.



6. **Mathematics:** Mathematics is an important part of data science. Mathematics includes the study of quantity, structure, space, and change. Good math skills are essential for data scientists.
7. **Machine learning:** Machine learning is the backbone of data science. Machine learning is training machines to function as the human brain. Data science uses a variety of machine learning algorithms to solve problems.

Data Science Tools

Below are some of the tools you need for data science.



- Data analysis tools: R, Python, Statistics, SAS, Jupyter, R Studio, MATLAB, Excel, RapidMiner.
- Data warehousing: ETL, SQL, Hadoop, Informatica / Talend, AWS Redshift
- Data visualization tools: R, Jupyter, Tableau, Cognos.
- Machine learning tools: Spark, Mahout, Azure ML Studio.

The Main Stages of the Data Science Life Cycle

Discovery: The first phase is discovery. This is all about asking the right questions. When you start a data science project, you need to determine basic requirements, priorities, and project budget. During this phase, you need to determine all the requirements for your project, such as: B. You can formulate business problems with people, technology, time, data, end goals, and first-level hypotheses.

Data preparation: Data preparation is also known as data tampering. In this phase, you need to perform the following tasks:

- Data cleaning
- Data reduction
- Data integration
- Data conversion,

After performing all the above tasks, you can easily use this data for the next process.

Model planning: In this phase, you need to determine different methods and techniques for establishing relationships between input variables. Apply

- SQL Analysis Services
- R
- SAS
- Python

Model building: This phase begins the model building process. Create datasets for training and testing purposes. We use various techniques such as association, classification, and clustering to build the model. Below are some common modeling tools.

- SAS Enterprise Miners
- Oh WEKA
- SPCS modeler
- Or MATLAB

Operationalization: This phase provides the final report of the project, along with briefings, code, and technical documentation. This phase gives you a clear picture of the full performance of your project and other small components before you fully deploy it.

Communicate Results: In this phase, you will see if you have achieved the goals you set in the initial phase. Communicate results and revenue to your business team.

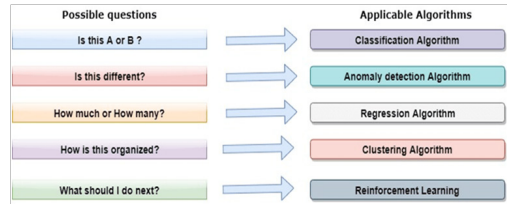
Data Science Application

- **Image and voice recognition:** Data science is currently used for image and voice recognition. When you upload a photo to Facebook and receive a suggestion to tag your friends. This auto-tagging proposal uses image recognition algorithms that are part of data science. If these devices respond via voice control, such as “OK Google, Siri, Cortana,” voice recognition algorithms make that possible.

- **Gaming World:** In the gaming world, the use of machine learning algorithms is increasing day by day. EA Sports, Sony and Nintendo make heavy use of data science to improve the user experience.
- **Internet Search:** Whenever you want to search something on the internet, use different types of search engines such as Google, Yahoo, Bing, Ask. All of these search engines use data science technology to improve the search experience. You can get search results in an instant.
- **Transportation:** The transportation industry is also using data science technology to develop self-driving cars. With self-driving cars, it's easy to reduce traffic accidents.
- **Healthcare:** In healthcare, data science has many advantages. Data science is used in tumor detection, drug discovery, medical image analysis, virtual medical bots, and more.
- **Recommender systems:** Most companies such as Amazon, Netflix, and Google Play use data science technology to improve the user experience with personalized recommendations. For example, if you search for something on Amazon and get a suggestion for a similar product, it's due to data science technology.
- **Risk detection:** The financial industry has always had problems with the risk of fraud and loss, but with the help of data science, this can be saved. Most financial companies are looking for data scientists to avoid all kinds of risks and losses while increasing customer satisfaction.

How can we use machine learning algorithms to solve data science problems?

Next, let's understand the most common types of problems in data science and how they are solved. Therefore, data science uses algorithms to solve problems. Below is a graphic representation of the algorithm that can be applied to the possible questions.

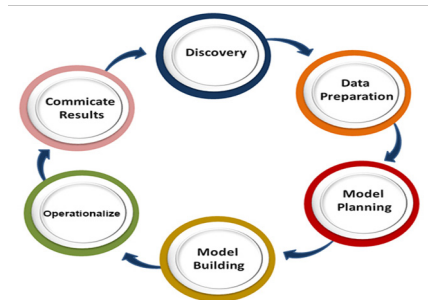


Is this wrong?

You can see this kind of question that belongs to different patterns and you need to find something strange among them. Such problems can be resolved using anomaly detection algorithms. How much is it and how many? Numerical values and other types of problems for numerical values such as B. Today's time, today's temperature can be solved by regression algorithms. How is this structured? Now, if you have a problem that requires you to deal with data cleanup, you can use a clustering algorithm to solve it. Clustering algorithms organize and group data based on features, colors, or other common characteristics.

Data Science Life Cycle

The data science life cycle is illustrated in the following figure.



The main stages of the data science life cycle are:

Discovery: The first phase is discovery. This is to ask the right questions. When you start a data science project, you need to determine basic requirements, priorities, and project budget. During this phase, you need to determine all the requirements for your project, such as You can formulate business problems with people, technology, time, data, end goals, and first-level hypotheses.

Data preparation: Data preparation is also known as data tampering. In this phase, you need to perform the following tasks:

Is this A or B?

Yes or No, you can see this kind of problem with only two fixed solutions like 1 or 0, may or may not. And this type of problems can be solved using classification algorithms.

- Data cleaning
- Data reduction
- Data integration
- Data conversion,

After performing all the above tasks, you can easily use this data for the next process.

Model planning: In this phase, you need to determine different methods and techniques for establishing relationships between input variables. Apply exploratory data analysis (EDA) using a variety of statistical formulas and visualization tools to understand the relationships between variables and see which data can provide us with information. Common tools for model planning are:

- SQL Analysis Services
- R
- SAS
- Python

Model building: This phase begins the model building process. Create datasets for training and testing purposes. We use various techniques such as association, classification, and clustering to build the model. Below are some common modeling tools.

- SAS Enterprise Miner
- WEKA
- SPCS Modeler
- MATLAB

Operationalization: This phase provides the final report of the project, along with briefings, code, and technical documentation. This phase gives you a clear picture of the full performance of your project and other small components before you fully deploy it.

Communicate Results: In this phase, you will see if you have achieved the goals you set in the initial phase. Communicate results and revenue to your business team.

Data Science Prerequisites

Non-Technical Requirements

- **Curiosity:** Learning data science requires curiosity. If you're curious and ask a variety of questions, it's easy to understand your business problem.
- **Critical Thinking:** Data scientists also need it so they can find some new ways to solve problems efficiently.

- **Communication skills:** Communication skills are of paramount importance to data scientists, as they need to communicate with the team after solving business problems.

Technical Prerequisite

- **Machine learning:** To understand data science, you need to understand the concepts of machine learning. Data science uses machine learning algorithms to solve a variety of problems.
- **Mathematical modeling:** Mathematical modeling is required to make rapid mathematical calculations and predictions from available data.
- **Statistics:** A basic understanding of statistics is required, such as: B. Mean, median, or standard deviation. You need to extract knowledge and get better results from your data.
- **Computer programming:** Data science requires knowledge of at least one programming language. R, Python, and Spark are the programming languages needed for data science.
- **Databases:** A deep understanding of databases like SQL is essential for data science to capture and manipulate data.

Conclusion

After weighing the strengths and weaknesses of data science, you can get a complete picture of this area. Data science is an area with many advantages, but it also has drawbacks. As a low-saturation, high-paying field that has revolutionized multiple life paths, it also has its own background given the vastness and interdisciplinary nature of the field. Data science is an ever-evolving field that can take years to master.

References

- Coursera, <https://www.coursera.org/browse/data-science>
- Great Learning Team. "What is Data Science & How Does It Work: A Complete Beginner's Guide." *Great Learning*, 2022, <https://www.mygreatlearning.com/blog/what-is-data-science>
- Insight Data Science Fellows Program*, <https://insightfellows.com/data-science>
- Sharma, Hemant. "What Is Data Science? A Beginner's Guide to Data Science." *Edureka*,

- <https://www.edureka.co/blog/what-is-data-science>
- Stedman, Craig. "What is Data Science? The Ultimate Guide." *Tech Target*, <https://www.techtarget.com/searchenterpriseai/definition/data-science>
- "BSc Data Science." *London School of Economics and Political Science*, <https://www.lse.ac.uk/study-at-lse/Undergraduate/Degree-programmes-2022/BSc-data-science>
- "Data Science (Msc)." *University of Southampton*, <https://www.southampton.ac.uk/courses/data-science-masters-msc>
- "Data Science." *IBM*, <https://www.ibm.com/cloud/learn/data-science-introduction>
- "Data Science." *Towards Data Science*, <https://towardsdatascience.com/data-science/home>
- "Data Science." *University of London*, <https://www.london.ac.uk/courses/data-science-msc>
- "Data Science Courses." *Ed X*, <https://www.edx.org/learn/data-science>
- "Data Science Tutorial." *W3 Schools*, <https://www.w3schools.com/datascience>
- "Post Graduate Program in Data Analytics and Data Science." *Skill Lync*, <https://skill-lync.com/computer-science-engineering-courses/pg-data-analytics-data-science>
- "What is Data Science?" *Data Robot*, <https://www.datarobot.com/wiki/data-science>
- "What is Data Science?" *Intellipaat*, <https://intellipaat.com/blog/what-is-data-science>

Author Details

K. Sankar, Assistant Professor of Computer Science, Lal Bahadur Shastri Govt. First Grade College, RT Nagar, Bangalore, India, **Email ID:** k.shankaranand@gmail.com.