



The Journal of Multidisciplinary Research (TJMDR)

Content Available at www.saap.org.in

ISSN: 2583-0317



Data science: applications and case studies

M Ramakrishna¹, K. V. M. Vara Kumar²¹Asst. Professor, Department of CSE, Adikavi Nannaya University Rajahmundry²Assistant Professor, Department of Mathematics, Adikavi Nannaya University, Rajamahendravaram, AP, India.*Received: 04 July 2022 Revised: 25 July 2022 Accepted: 22 Aug 2022*

Abstract

Data science is the activity of analysing vast amounts of unorganised and organised raw data to find patterns and draw conclusions that can be put to use. Data science is an interdisciplinary field, and the foundations include, inference, computer science, predictive analytics, the creation of machine learning techniques, and new tools for extracting information from large data sets. Data science is still in its beginning stages of development, but it is already producing professionals with specific and relevant skills that set them apart from those in the computing, digital, and statistical sciences.

Keywords: Data science, statistics.

This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. Copyright © 2022 Author(s) retain the copyright of this article.



*Corresponding Author

M Ramakrishna

Produced and Published by

South Asian Academic Publications

1. Introduction:

To examine massive amounts of organised and unorganised big data to look for trends thanks to data science. In essence, this enables businesses to strengthen their competitive advantage, control expenses, find new market opportunities, and improve productivity.

To define data science and improve data science project management, start with its lifecycle. Data acquisition, occasionally data extraction, and data entry are all steps in the data science pipeline workflow's initial stage, known as "capture." The maintenance phase follows, and it entails data processing, data cleaning, and data staging etc.

The next step is data processing, which is a foundational concept in data science. Data analysts differentiate themselves from data engineers during the data exploration and analysis stages. The steps that produce valuable data are included in this phase, including data mining, data classification and clustering, data

modelling, and summarising insights obtained from the data.

Data analysis is the subsequent step, which is also very important. Here, data analysts perform text mining, regression, qualitative analysis, predictive analysis, and exploratory and confirmatory work. While data science is carried out effectively, there isn't any such thing as cookie-cutter data science. The data scientist shares observations in the last phase. This entails data reporting, data visualisation, and the use of various business intelligence technologies in order to help organisations, decision-makers, and other people make better decisions.

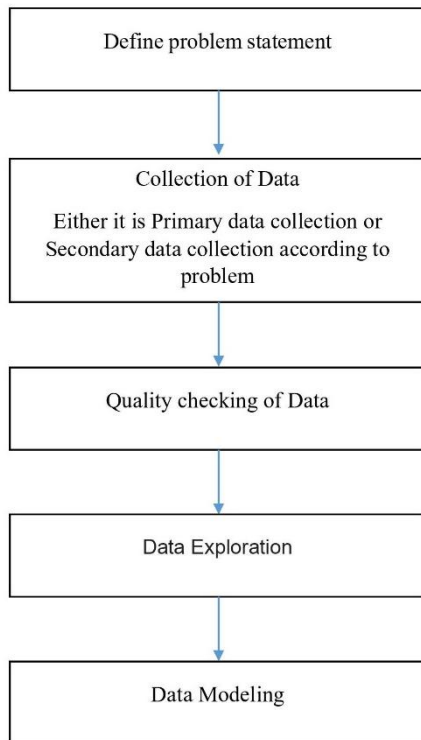
Data analysis requires a number of prerequisites, including developing practical questions, organising experiments, making systematic observations, interpreting and analysing data, drawing findings, communicating, and organising and carrying out a thorough inquiry.

2. Data Science Life Cycle:

Several data science life cycles concentrate merely on the data simulation, and analysis stages. Some are more in-depth, beginning with business insight and ending with implementation. And the one we'll go over is significantly more comprehensive, including operations. It also places a greater emphasis on

mobility than other life cycles.

The lifecycle of Data Science has five steps as shown in flow chart:



Step1: Define Problem Statement

Establishing a precise problem statement, which is a synopsis of the problem you're going to tackle, is the fundamental and most important stage in data science. Yet why do we require a precise problem statement? Furthermore, all of our efforts and activities, aside from identifying the problem statement, are aimed at resolving it. The problem statement must be unambiguous, have a main objective, and be verifiable. This can be accomplished by asking the appropriate questions.

Step2: Data Collection

A systematic approach to gathering relevant information from a variety of sources is known as data collection. You must gather data that will assist you in solving the situation. The data collection approach is widely categorised into two types based on the problem statement, namely primary data collection and secondary data collection.

Primary data collection is the collection of new data, when you have a one-of-a-kind situation but no ongoing statistics can be accessed via the internet, official censuses, periodicals, in news articles and other places. This procedure takes less time than the primary.

Step3: Quality Checking of Data

The need to guarantee the integrity of the information used for evaluation and explanation is sometimes disregarded by data analysts. After acquiring data, most individuals begin analysing it. They consistently did not perform data reality testing. If the data is of poor quality, the results can be misleading.

Step4: Data Exploration

Data exploration is the process of efficiently obtaining knowledge from data, even when we are unaware of the information we are searching for. Data exploration is intended for static applications with a large number of workload assumptions. Certainly, managing an employee or inventory database is a very different environment from searching for interesting trends in a scientific database [6].

Step 5: Data Modelling

Modelling entails outlining each phase and collecting the techniques needed to attain the solution. You must document the flow of the calculations, which are just adding modelling elements to the solutions. The essential issue is how the calculations are performed. ML and Statistics provide a variety of strategies from which to choose, depending on your needs.

3. ApplicationsofDataScience:

The role of Data Science Applications hasn't evolved overnight. Thanks to faster computing and cheaper storage, we can now predict outcomes in minutes, what could take several human hours to process.

- **Finance Sector**

Finance was one of the first industries to use data science. Every year, businesses were tired of poor loans and deficits. However, they have a plethora of information that was obtained during the first round of documentation while granting loans. They chose to bring in data analysts to help them recover from their deficits.

Banking firms have learned to split and control data over the years by using consumer profiling, historical spending, and other critical information to estimate volatility and default possibilities. Furthermore, it enabled them to promote their banking products based on the purchasing power of their clients.

- **Health Care Sector**

The burgeoning requirement for clinical governance to manage the stream of clinical data supporting evidence-based medicine and enhancing the effectiveness of health services is recognized to have a viable solution in

data analysis. In this sector, big data analysis reduces expenses, facilitates better management of diseases caused by a sedentary lifestyle, simplifies bureaucratic challenges, and enhances interactions between patients and doctors.

Investigations in occupational and preventative medicine frequently start with data gathering from the customer's regular activities. This necessitates practical measuring solutions using physiological, psychological, physical, and occasionally emotional criteria [12].

- **Speech Recognition**

Even if you are unable to compose a message, using the speech recognition capability would allow you to continue living your life. Google Voice, Siri, and Cortina are some of the best examples of speech recognition products. The message can be translated to text by just speaking it out loud.

- **Advanced Image Recognition**

When you post a photograph of yourself and your pics on social media, others start suggesting who you should tag. An algorithm for face recognition is used in this automatic tag recommendation feature. Recently, an update outlining the extra work they've achieved in this area, specifically highlighting their improvements in image recognition accuracy and returns relevant search results.

- **Gaming**

Modern video games are developed using ML algorithms, which get better as the player advances to a higher level. The computer opponent in motion games likewise studies your prior moves and adjusts its strategy accordingly. With the use of data science, most gaming apps have elevated the gaming experience.

- **Genetics and genome**

Through the use of data science applications, high levels of therapeutic customization are also made possible by genetic and genomic research. Understanding how our DNA affects human healthcare and identifying individualised biologic links between genetics, diseases, and treatment response are the main objectives. By combining multiple types of data with genome sequencing using big data tools, pharmaceutical research can gain a greater understanding of the role that genes play in how certain diseases and treatments affect people.

4. Advantages and Disadvantages of Data Science:

The vast subject of data science offers its own set of benefits and restrictions. Therefore, we shall assess the benefits and drawbacks of data science here. The

following are some of the many advantages of data science.

4.1 Advantages of data science:

- **Sought-After**

Data science is highly sought-after. The chances for prospective employees are numerous. Data Science is currently the fastest-growing occupation, making it an incredibly employable career field.

- **Numerous opportunities**

The necessary skill set needed to become a full-fledged data analyst is possessed by a very small number of people. As a result, data science is less developed than other IT industries. As a result, the subject of data science is quite diverse and offers many options. Data analysts are in abundant supply, yet there is a shortage of them.

- **Nomoretious activities**

Automation of redundant work has been made possible by data science across several sectors. Corporations are training computers to carry out repetitive activities by leveraging past data. This has made the difficult tasks that humans once performed simpler.

- **Improves product capabilities**

ML is used in data science, which has allowed businesses to improve their goods and make them more suited to the needs of their customers. Recommendation systems, such as those employed by e-commerce websites, offer users individualized insights based on their prior purchases. This has made it possible for computers to comprehend human behavior and make informed decisions.

- **Datasciencecansaveslives**

Data science has significantly enhanced the health sector. ML has made it simpler to find malignancies in their initial stages. A lot of other sectors in the field of health care are now adopting data science to benefit their customers.

4.2 Disadvantages of datascience:

Although a job in data science can be quite profitable, there are a number of drawbacks. We also need to be aware of the drawbacks of data science in order to fully comprehend it. Here are a few of them:

- **Data analysis is a hazy field**

The term "data science" is quite broad and lacks a clear definition. The definition of a data analyst is incredibly elusive, despite the fact that it has evolved into a term. The particular responsibilities of a data analyst vary depending on the industry in which the organisation operates. Few detractors have referred to data science as nothing more than a rebranded version of analytics, despite the fact that others have referred to it as the fourth paradigm of science.

- **Almost impossible to become an expert in data science**
Data science, which combines a variety of disciplines, has its roots in mathematics, statistics, and computer science. Being a true expert in every discipline and a maestro in each is quite unlikely. Despite numerous virtual classrooms attempting to close the skill gap that the data science sector is experiencing, the vastness of the field makes it impossible to become proficient in it. It's possible that a person with a background in statistics won't be able to quickly learn computer science and become an expert data scientist. As a result, it is a dynamic, ever-changing area that necessitates continual study of the numerous applications of data science.

- **Significant technical skills are required**
The reliance of Data Science on technical skills is another drawback. It will be challenging for someone without the necessary foreknowledge to resolve Data Science problems, even if they have a strong foundation in computing and statistics. This also applies to the other way around.

- **Concern with data privacy**

Data is the fuel for numerous industries. Data analysts support businesses in their decision-making using data. The procedure, however, may violate clients' privacy with the data used. The firm usually can see the personal information of customers, which occasionally results in data leaks because of security vulnerabilities. Several sectors have been troubled by the moral dilemmas around the privacy and ethical preservation of data.

- **Random data could produce freak occurrences**
To aid in judgment call, a data analyst examines the data and formulates thoughtful projections. The data given is frequently arbitrary and fails to produce the desired results. This could also fail as a result of inadequate resource management.

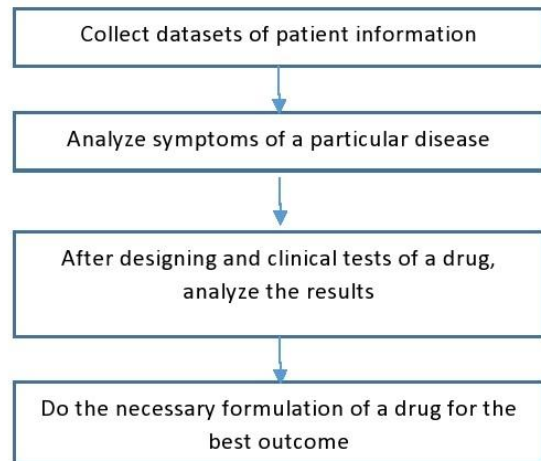
5. DataScienceCaseStudies:

Here are some of the most well-known data science case

studies that will explain how data science is applied across many industries. The significance of data science in various fields.

- **Pharmaceutical industries**

With the advancement of cloud-based technology and data analytics, it is now simpler to examine huge amounts of patient data. Medicine has been transformed in the pharmaceutical industry by artificial intelligence and data science. It is challenging for doctors to stay current on possible treatments given the daily arrival of



new pharmaceuticals. Additionally, a more generalised diagnostic and therapeutic approach finds it challenging to enter a highly competitive market. However, pharmaceutical firms can now more easily gain a competitive edge over the market because of developments in analytics and the development of statistical approaches that are pipelined in parallel.

- **Predictive Modelling for Maintaining Oil and Gas supply**

Machinery breakdowns, which frequently result from petroleum wells' inefficiencies and exceptional performance, are a significant issue for the crude oil and gas industries. By implementing an effective plan that supports preventive modelling, operators can be informed of maintenance windows as well as critical suspension phases. This will increase production of oil and stop additional losses. Data analysts can utilise process monitoring techniques to optimise the slightly elevated machinery used in the production and refinement of oil products. Our machine learning model may be trained using historical information that is continuously streamed using the sensor information that is retrieved from sensors.

This artificial intelligence model will forecast the failure of machine components and alert the operators

to the need for prompt repair to prevent oil losses. A data analyst tasked with developing the approach will aid in avoiding dangers and forecast machine malfunctions, alerting the operators to take preventative measures. Data analysts can employ proactive maintenance strategies to optimize high-value equipment used in the production and refinement of oil products.

- **Educational sector**

The manner that students are engaged with teachers and are graded on their performance has also altered as a result of data science. Teachers can utilize data science to examine student feedback and use it to enhance their instruction.

However, a robust expansion in data science education that will irrevocably influence future undergraduate students, regardless of its final prospective discipline classification. Data science instruction is actually expected to become a standard part of the undergraduate experience, driven by rising student interest and corporate demand. As the importance of data skills is further understood, there will be a rise in the number of students majoring, minoring, getting certificates, or merely taking courses in data science.

- **Biotechnology**

A, T, C, and G are the four building pieces that make up each human gene. These four basic pieces can be combined in 3 billion different ways, which determines how we look and behave. While there are both hereditary and lifestyle-related abnormalities, the results of both can lead to long-term diseases.

- **HotelRecommendationSystem**

Usually, a hotel recommendation system uses reinforcement learning, which bases recommendations on reviews provided by other consumers that fall within the same exact category as the individual seeking a product.

Everybody plans vacations, and securing lodging is the primary step in the process. Numerous websites offer advice on which hotel would be good for our excursion. A hotel suggestion system seeks to predict which resort among the resorts a visitor will most likely select. Therefore, to create this kind of system that will assist the user in choosing the best resort out of all the resorts. Customer feedback will assist us with this.

For instance, if you plan to travel for business, the hotel recommendation system should display the accommodations that previous clients have rated as the finest for business travel. Establishing a recommendation system based on consumer feedback

and ratings is thus also our strategy. So, create a hotel recommendation system using the ratings and reviews provided by users who fall into the same category as the user.

6. Future Scope in Data Science:

Numerous factors are influencing the direction of data science as the field progresses. For starters, more data science functions in the life cycle are likely to be automated. With more big data procedures being automated, so much data will be available to more individuals in far more industries, and AI and machine learning will advance rapidly.

Another trend could be the availability of more data science resources for a broader audience. Data scientists often have narrow skill sets. However, there is an increasing demand for both who can properly accomplish data science activities and professionals to manage AI and ML projects in particular. This expansion is fuelling a tendency in science communication.

The tension between privacy rights and the need to regulate and hold people accountable is another trend that will undoubtedly influence the development of data science. Data science has the potential to make ML algorithms and the method by which we train AIs significantly more public, allowing for regulatory monitoring.

This is extremely prevalent in specific business areas where a web address or industry experience is required. More sophisticated processes, similar to other areas of science and technology, may be designated for data analysts with more specialised skills, but less specialised activities will become more accessible. It ought to be fascinating to observe what additional areas of using data science might broaden as automation evolves.

7. Conclusion:

The field of data science is well into its early phases of development; it is becoming a self-sustaining discipline and creating individuals with unique and complementary skills compared to those in the computer, information, and statistical sciences. Data science includes a number of techniques to deal with reproducibility. It has generated significant outcomes in a number of areas.

References:

1. An introduction to data science by Jeffrey S.Saltz, JeffreyM.Stanton
2. Wing, Jeannette M. "The

- data life cycle." *Harvard Data Science Review* 1.1 (2019): 6.
3. Baesens, Bart. *Analytics in a big data world: The essential guide to data science and its applications*. John Wiley & Sons, 2014.
 4. Hox, Joop J., and Hennie R. Boeije. "Data collection, primary versus secondary." (2005): 593-599.
 5. Idreos, Stratos, Olga Papaemmanouil, and Surajit Chaudhuri. "Overview of data exploration techniques." *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. 2015.
 6. Wang, Yichuan, et al. "An integrated big data analytics-enabled transformation model: Application to health care." *Information & Management* 55.1 (2018): 64-79.
 7. Bates, David W., et al. "Big data in health care: using analytics to identify and manage high-risk and high-cost patients." *Health affairs* 33.7 (2014): 1123-1131.
 8. Murdoch, Travis B., and Allan S. Detsky. "The inevitable application of big data to health care." *Jama* 309.13 (2013): 1351-1352.
 9. Manyika, James, et al. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute, 2011.
 10. Hripcsak, George, et al. "Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers." *MEDINFO 2015: eHealth-enabled Health*. IOS Press, 2015. 574-578.
 11. Hripcsak, George, et al. "Drawing reproducible conclusions from observational clinical data with OHDSI." *Yearbook of medical informatics* 30.01 (2021): 283-289.
 12. Neubert, Sebastian, et al. "Multi-sensor-fusion approach for a data-science-oriented preventive health management system: concept and development of a decentralized data collection approach for heterogeneous data sources." *International Journal of Telemedicine and Applications* 2019 (2019).
 13. Kaur, Parneet, Manpreet Singh, and Gurpreet Singh Josan. "Classification and prediction based data mining algorithms to predict slow learners in education sector." *Procedia Computer Science* 57 (2015): 500-508.
 14. Ali, Mohd Maqsood. "Role of data mining in education sector." *International Journal of Computer Science and Mobile Computing* 2.4 (2013): 374-383.
 15. Hassani, Hossein, et al. "Deep learning and implementations in banking." *Annals of Data Science* 7.3 (2020): 433-446.
 16. Kou, Gang, et al. "Machine learning methods for systemic risk analysis in financial sectors." *Technological and Economic Development of Economy* 25.5 (2019): 716-742.
 17. Kou, Gang, et al. "Machine learning methods for systemic risk analysis in financial sectors." *Technological and Economic Development of Economy* 25.5 (2019): 716-742.
 18. Bangert, Patrick, ed. *Machine learning and data science in the oil and gas industry: Best practices, tools, and case studies*. Gulf Professional Publishing, 2021.
 19. Sayers, Eric W., et al. "Database resources of the national center for biotechnology information." *Nucleic acids research* 49.D1 (2021): D10.
 20. Saputro, Pujo Hari, and Herlino Nanang. "Exploratory data analysis & booking cancellation prediction on hotel booking demands datasets." *Journal of Applied Data Sciences* 2.1 (2021): 40-56.