SKILLS ASSOCIATED WITH DATA SCIENCE

Shamli Prasad*

Abstract:
Data science is a field that is changing dramatically over time. One should be aware that the application of its components can take a long time to complete. It is a good mix of data extraction, analysis, and modern technologies. In today's world, the amount of unrecognized data on the open platform is quite large. A data scientist must be aware of the skills used to handle such an amount of data. To clarify, let's take an example to understand this scenario. Healthcare is a vast field with multiple data logs and categories of areas of information. Areas like medicinal information, patient records, diseases history and occurrence, and many more. Analyzation of this data with utmost accuracy and speed with modern requirements can help the healthcare sector.

Data science is an essential tool in this modern age. Companies are constantly improving their technical skills to meet the ever-changing competitive market. It has become a weapon of the impending technological age by setting market standards and stimulating global investors. However, the market is constantly growing, and it becomes difficult to determine the latest tools and skills required by businesses. Therefore, a data scientist needs to match up with such an advancing field by evolving his skill set and keeping up with trends of modern requirements.

Here we have discussed the requirements and trends needed by an aspiring data scientist. Industries have marked must have standards that you need to familiarize yourself with to get a job. Although, you must also familiarize yourselves with data science and its working nature.

*Chandigarh University, Punjab, India Email: shamliprasad.cs17@gmail.com

***Corresponding Author:** Shamli Prasad
* Chandigarh University, Punjab, India Email: shamliprasad.cs17@gmail.com

DOI: 10.48047/ecb/2023.12.si10.00472

1. Introduction

Data science is data-driven and identifies questions that re quire answers in a given situation. It also determines the cor rect analysis to find the solution to a given problem. Data Science is an essential part of finding answers from multi ple sources, organizing them into an organized fact sheet for favorable outcomes on business decisions. Therefore, pre dictions and decisions by data scientists using different data models and theories use either a predictive analytic approach or a prescriptive analytic approach. It is a follow-up approach that focuses on critical past and current situations. To be a data scientist, you must master advanced data science skills.

In addition, data scientists use the latest technology to achieve viable solutions. They merely present the data com pared to the raw data available. The raw data undergoes var ious stages before being used for analysis. This filtration and processing demand highly skilled human resources, data sci entists. Therefore, he should match up with modern skills to satisfy the industry's requirements and job essentials.

2 Skills associated with Data Science

2.1 Python

Python is a high-level interpreted generic programming lan guage. Its design philosophy emphasizes the code readability by the usage of meaningful indentation. The first of the many advantages of Python in data science is its simplicity. While some data scientists have a background in computer science or know other programming languages, many have a back ground in statistics, math, or other technical fields and may not have as much experience in code when they do. Python's syntax is easy to follow and write, making it a straightforward programming language to start and learn quickly.

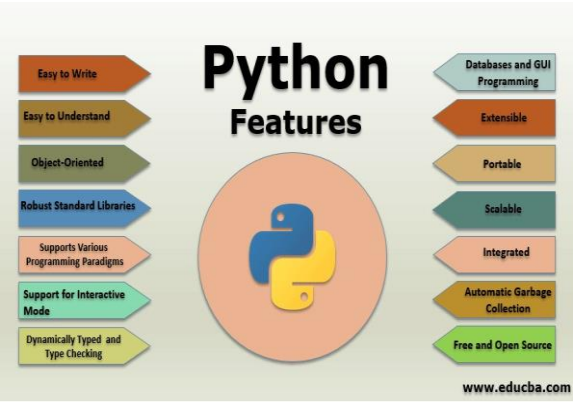

FIGURE 1. Features of Python

2.2 Analytic

Analysis plays a vital role in solving a problem with proper research. However, technologies are not directly classified for analysis. It depends on the basics of complete visualiza tion and the requirements of the situation in question. Statis tics is a course to get the best analytical process. It will help define a set of libraries focused on data visualization. Statis tics are a necessary basis for going into detail and extracting information from a given set of data. In addition, it helps quantify uncertain data sets and analyze data set behavior. It is, therefore, necessary to acquire statistical knowledge for a data scientist.

FIGURE 2. Data Analytic Process

2.3 Data wrangling

In Data Science, the data we deal with can be complex. Therefore, it is necessary to decompose the problem of com plex data into simple datasets. Understanding how to handle errors in a data set is essential. Using data management, you can process error-free data for analysis and troubleshooting.

FIGURE 3. Data Wrangling

2.4 Machine learning

Many responsibilities rest on the shoulders of data scientists. Charged with the crucial task of identifying and solving a business problem, they must convert it into machine learning activities. Machine learning skills feed relevant data-driven algorithms and models to solve the task at hand.

Machine learning processes real-time data that uses different data driven models to predict data patterns and produce precise results by feeding data sets.

2.5 Linear algebra and Calculus

Linear Algebra is useful in data science. It is the most es sential mathematical skill used in machine learning. Most machine learning models can be drawn in matrix form. Also, it is used in data preprocessing, transformation, and model evaluation. Besides, learning algebra and calculus helps in data visualization as well as in data analysis. It simplifies the raw data and even calculates the future possibilities for the same.

2.6 Dev-ops

DevOps is a data science skill that is considered unimportant. Data models, put in place to solve problems, must sometimes respond to the virtual environment. Therefore, DevOps be comes a crucial part of data science management. In addi tion, these skills can help solve complex theories in program ming to obtain optimal solutions. You may have difficulty implementing internal DevOps if you don't know how to set things up. Data science needs to be familiar with terminals and CLIs.

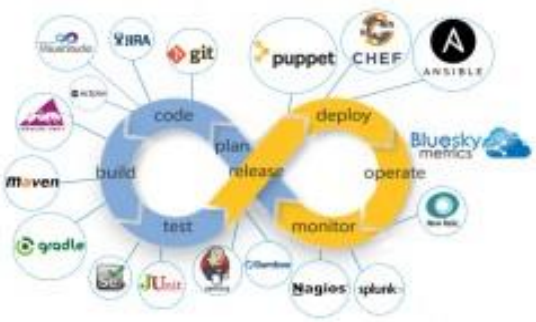

FIGURE 4. Dev-Ops Tools

2.7 Storytelling skills

Storytelling is a vital data scientist skill that you must learn. It becomes necessary to improve storytelling skills to develop datasets. You are the problem solver in your business and everyone depends on you to find clear solutions to complex problems. Convert quantitative solutions into language ev eryone in your business can understand. Your responsibili ties lead you to translate the results into a language used by the company. Also, you have to tell data stories. The data in your hand is your character and you have to create the texture around it. Make sure everyone understands your language. Therefore, storytelling can help you communicate with your colleagues and make them understand the data.

2.8 Collaboration

You can fulfill your dream of becoming a data scientist if you learn to channel your knowledge to increase the speed of exit.

You have to grow your business. Doing heavy tasks alone be comes difficult. Collaborating with your team helps in solv ing complex problems. Increase the pace of your work while keeping the quality at the top. Therefore, human resources can help overcome business challenges while meeting dead lines.

2.9 Performing data experiments

Experiment with datasets and create new datasets. Data sci entists can access unknown data sources at any time. Hence, it becomes necessary to practice and experiment on the data. The world is full of data. As a data scientist, your ultimate goal is to manage and solve real-world entities. Every day we are faced with multiple unknown problems, and as a data scientist, you are faced with the challenge of solving these problems. The experience you gain by experimenting will only help you in your professional career.

2.10 Domain Knowledge

Data science is becoming a key element in every industry. As a data scientist, you need to attain knowledge about your work. You should be aware of problem areas such as manu facturing, fashion, automotive, real estate, education, sports, etc. Gaining knowledge about the field can help with long term data science projects. Examining the available data sources can help you familiarize yourself with the business niche. Common mistakes caused by data scientists are that they rush into collecting and analyzing data rather than un derstanding the demands of the problem in question. Under standing the business framing throughout the cycle is essen tial to achieve smooth management and steady progression.

3 Critical Review of future aspects

Data Science is an evolving field which deals with statistics and programming. As internet availability is growing, data on the net is drastically increasing. This increase cannot overcome the argument that data scientists will fade away in the future due to the highly technical approach and automated tasks by predesigned software.

Although, there are fields of data science where data scientists can involve without worrying about future aspects. With skills like machine learning, data wrangling, python and more, a data scientist

can evolve their skills to handle highly statistical and complex mathematical problems.

In a typical career, there are various dissolving factors. Though some skills like linear understanding, machine learning, programming with python and data experimentation, fields like machine learning engineers and data scientists with graph and adaptive science can progress much ahead with these skills.

4 Conclusion

These high-level skills can help you achieve your goal as a data scientist. As the tech industry grows, data science will become mandatory in all fields in the future. It can seem un certain to predict the future of progress as it evolves rapidly. To the best of my knowledge, many different changes are happening in this field. Ultimately, the world today belongs to data science. More and more data will provide decision making opportunities. It will change the way we see a real world entity. Therefore, a data scientist must be highly skilled and aware of new needs and advancement

5 References

1. Prüfer, Jens, and Patricia Prüfer. "Data science for en trepreneurship research: studying demand dynamics for en trepreneurial skills in the Netherlands." Small Business Eco nomics 55.3 (2020): 651-672.
2. Oliphant, Travis E. "Python for scientific computing." Computing in science engineering 9.3 (2007): 10-20.
3. Terrizzano, I. G., Schwarz, P. M., Roth, M., Colino, J. E. (2015, January). Data Wrangling: The Challenging Journey from the Wild to the Lake. In CIDR.
4. Li, W., Zhang, Y., Sun, Y., Wang, W., Li, M., Zhang, W., Lin, X. (2019). Approximate nearest neighbor search on high dimensional data—experiments, analyses, and improvement. IEEE Transactions on Knowledge and Data Engineering, 32(8), 1475-1488.
5. Ebert, C., Gallardo, G., Hernantes, J., Serrano, N. (2016). De vOps. Ieee Software, 33(3), 94-100.