Modern Era of Python: In Data Science

TusharVarun[1], MS Himanshu Gupta[2]
[1]M.sc Data Science, Chandigarh University, India.
[2]Assistant Professor,Department of Computer Science,Chandigarh University, India.
[1] tusharvarun989@gmail.com, [2] himanshu.e11818@cumail.in

Abstract

"Python is a high-level, open-source, interpreted language that offers a fantastic approach to object-oriented programming". Data scientists utilize it as one of the best languages for a variety of projects and applications. Python has excellent capabilities for working with mathematical, statistical, and scientific functions. It offers excellent libraries for dealing with applications of data science. Because of its simplicity and ease of use, Python is one of the most popular programming languages in the scientific and research sectors. People without technical backgrounds may easily learn how to use it because of this. Additionally, it is better suited for rapid prototyping.

Keywords Data Analysis ,DataScience ,Machine Learning ,Deep Learning ,Artificial intelligence ,Artificial Neural Network, Convolutional Neural Network , Data Visualization, Data Mining, Natural Language Processing, Recurrent Neural Network

DOINumber:10.14704/nq.2022.20.10.NQ55808 **NeuroQuantology 2022; 20(10): 8235-8240**

INTRODUCTION

Python is a high-powered, demanding, freely available, interpreted programming language that offers a brilliant approach to object-oriented programming. It is one of the most popular computer languages in the field of DS for different projects and its functioning**[1].** It has tonne of features for handling arithmetic, statistics, and scientific functions. It has brilliant libraries for data science projects.

It's famous in the fields of research for its ease of use and friendly syntax, which makes it understandable even for people who have no technical background, and it's also better for swift prototyping.

It also seems like a better option than other programming languages for ML scientists when it comes to application fields and its libraries. Developers select Java for several areas such as fraud detection algorithms and network security, while Python was selected for applications like NLP and sentiment analysis because it provides a massive pair of libraries that help solve hard business problems rapidly and also create strong systems and data applications.

Its demand in the area of DS has surpassed all previous records and it is also utilised by 66 % approx of around 1950 plus members in a poll released in the month of May and year 2018 by the authoritative portal KDNuggets **[1],** under the category "Top Analytics, DS, ML Tools," compared to 53 % approx for RapidMiner and 49 % approx for R, its 2nd greatest competitors. In practise, it has become the first computer language of choice for the data scientist community during the previous three years, with R coming in second.

*e*ISSN1303-5150

LITERATURE REVIEW TABLE

Author(Name)	Research
Stančin, I., & Jović, A. (2019, May)	The author in [1] highlighted the features of DM as well as large DA packages. There is only one paper which defines the pros and cons of many libraries. It also shows more than 18 libraries, divided into six categories.
Murtagh, F., & Devlin, K. (2018)	The author in [2] covers the rapid growth of DS, and the information is collected from online websites, surveys, and job reports, which is quite good.
Cao, L. (2017)	The author in [3] highlighted the overview of DS and collected the information from the survey and tutorial of DS.
Kudande, D. V., & Shah, B. B.	The author in [4] talks about how DS is made and also covers some other major learnings. Information gathered from handbooks, websites, and other articles
Siva Jyothi, P. N., & Yamaganti, R. (2019)	The author in [5] paper describes the importance of the Python language for building IoT apps and also discusses some other factors like how Python is connected to ML, DL, and DS. For this article, the author collected the information from many websites and other published papers that are based on modern technologies.
Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020)	The author in [6] highlighted the algorithm of scientific computing in Python, which covers the uses of scipy and information taken for this article from github and other papers.
Butwall, M., Ranka, P., & Shah, S	The author in [7] covers all the important methods, tools, and libraries which are based on the data science area, and the author covers the ML concept with its essential libraries. The information collected for this article came from many reviews and websites.
Saabith AS, Vinothraj T, Fareez MM	The author in [8] points out the Python libraries and Ides for DS.

PYTHON BASIC FEATURES:

Python is a high-powered, interpreted, interactive, open-source programming language that may be used for a variety of tasks and activities. It was created by Guido van Rossum between 1985 and 1990. Its source code is released under the GNU General Public License, similar to Perl (GPL).

1) ***Interpreted Python***- *The interpreter handles Python processing at runtime. Before you start using your software, you do not need to assemble it. This is a scripting language similar to PERL and PHP.*
2) ***Interactive Python***- *At a Python prompt, you can truly sit and*

interconnect directly with the interpreter to write your programs.

DATA SCIENCE FORMATION:

Here we will discuss the formation of data science with the help of important figures. Figure 1.1 define the some fundamental areas related to data science**[2].**

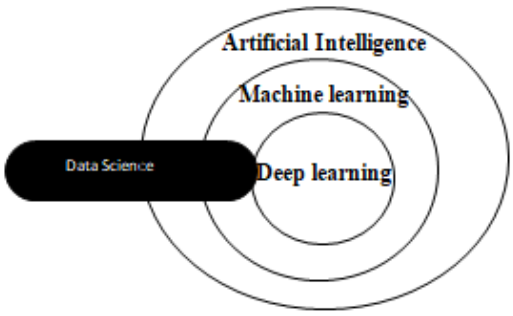


Figure 1.1: Fundamental areas and data science formation.

Data Science consist of ***Artificial Intelligence, Machine Learning, Deep Learning .***

1) ***AI :*** *It is a pair of stats and technology that allows advanced machines like human robots to be created****[3].*** *In this context, AI is a method of teaching machines to act like humans. AI examples include self-driving cars, Amazon Alexa, and Android apps such as Cortana and Google Maps****.***
2) ***ML :*** *It is the part of AI which explores and analyses data using statistical tools. The three types of ML are as follows****[4]:***
   a) ***supervised learning****: This type of learning has tagged data for classification and regression.*
   b) ***unsupervised learning:*** *Untagged data is present & algorithms such as clustering, kNN, and k-means.*
   c) ***semi supervised learning:*** *It uses a pair of labelled and unlabelled data. Take, for instance, speech analysis and content classification. Its algorithms, such as generative models and also those based on graphs,*
3) ***DL****: It is a branch of ML that uses architectures such as multi-neural-networks to simulate the brain. It can also be classified into 3 categories****[4]:***
   a) ANN: *Numbers-based data set.*
   b) CNN**:** *Images-based data set.*
   c) RNN: *Time series data is used to create this dataset.*

PYTHON LIBRARIES FOR DATA SCIENCE:

It has some important libraries which have an essential role in the field of DS, such as**[5]:**

- *Numpy*
- *Matplotlib*
- *Scipy*

Numpy:

A Python library for working with arrays is called NumPy. Additionally, it has matrices, fourier transforms, and functions for working in the field of linear algebra. Travis Oliphant invented NumPy in 2005. You may use it without restriction because it is an open source project. Among the features of NumPy are:

- *ndarray is a quick and efficient multidimensional array object.*
- *Tools for reading and writing array-based datasets to disk*

NumPy's principal purpose for DA is as a storage for data to be exchanged between algorithms and libraries, in addition to the fast array-processing capabilities it offers to Python

Matplotlib:

An easily available package for making a plot and other 2D DV is matplotlib. It was founded by John D. Hunter and is currently handled by a large development team. It's made for plot making, which is good for publication. While Python programmers have access to various visualisation libraries, matplotlib is the most commonly used and, as a result, has a high level of integration with the rest of the ecosystem.

*e*ISSN1303-5150

As a default visualisation tool, I believe it is a safe bet.

Scipy:

SciPy is a free-source library for tackling math, engineering, and technology problems. It works for the manipulation and visualisation of data. The NumPy extension is used to build SciPy. "Sigh Pi" is how SciPy is pronounced. Since its inception in 2001, it has established a de facto standard for exploiting techniques in Python, with over 580 distinct code participants, thousands of dependent packages, over 90,000 dependent repositories, and millions of instals each year [6].

PYTHON IDE FOR DATA SCIENCE:

It provides a variety of IDEs for various uses. However, there are certain editors that can be utilised in DS**[7].**

- ***Spyder:*** *Spyder was created exclusively for DS, unlike most other IDEs on the internet**[7].** This editor includes tools such as syntax highlighting, code completion, and variable editing, all of which may be edited via a graphical user interface (GUI).*
- ***PyCharm:*** *PyCharm is an IDE designed by JetBrains, the company behind one of the most well-known Java IDEs**[7].** PyCharm includes NumPy and Matplotlib in its tools and libraries, allowing you to work with array viewers and plots. PyCharm is a strong alternative for web development because it promotes JavaScript, XML, CSS, YAML, and other languages in addition to Python. PyCharm, like other IDEs, includes useful tools including a compiler, error highlighting, and a debugger with a graphical interface, in addition to Git. You may also personalise your IDE by selecting from a variety of themes, colours, and key bindings. Plugins can also be used to extend PyCharm's capabilities.*
- ***Thonny:*** *Thonny is the next IDE, which is designed to make programming easy. It's a piece of software created at Tartu University. Thonny includes code completion and syntax errors, as well as a simple debugger that allows you to compile your programme piecemeal. Beginners will appreciate this because it can be learned through statements and expressions. A new window appears while editing a function, with local variables and code displayed separately from your input.*
- ***Atom:*** *Github has created an open-source text editor. Although Atom is available for a variety of programming languages, including Scala, HTML, GO, SQL, CSS, and others, it has certain unique capabilities which make it an excellent choice for Python developers.*

SOME APPLICATION OF DATA SCIENCE:

*DS has established a solid footing in a number of sectors, including healthcare, finance, manufacturing, and transportation. It has several uses and a wide range of applications. The following are some examples of data science applications***[8]***:*

1) ***Banking:*** One of the most common uses of DS is in banking. Banks are now able to compete because of big data and DS. Banks can effectively manage their resources with the use of DS, and they can also make better judgments through fraud detection, customer data management, risk modelling, real-time predictive analytics, customer segmentation, etc. In order to track their client base, banks also evaluate the lifetime value of each individual customer. It provides them with numerous projections about what the business bank will learn from its clients. Banks give businesses the ability to identify scams involving credit cards, insurance, and accounting. Additionally, banks are able to examine the cycles and patterns of consumer

investment in order to provide offers that are suitable for you. Additionally, banks have access to data science risk modelling tools that allow them to evaluate their entire performance. Banks can create individualised marketing that meets the demands of their customers thanks to data science. Banks utilise machine learning algorithms to enhance their analytics approach in real-time and predictive analytics. Banks also employ real-time data to identify the underlying issues that hinder their performance.

2) ***Manufacturing:*** Data scientists are today's equivalent of manufacturing employees. In the manufacturing sector, data scientists have therefore taken on a crucial role. Data science is widely applied in the manufacturing sector for production optimization, cost-cutting, and profit-boosting purposes. Additionally, the inclusion of technology like the Internet of Things (IoT) has made it possible for businesses to foresee future issues, monitor systems, and analyse the constant stream of data. Industry may also monitor its energy expenditures and improve output with the help of Data scientists may assist enterprises in making better judgments and raising the calibre of their products by thoroughly analysing client feedback. Automation is a key component in several businesses. With the use of historical and real-time data, industries are able to create autonomous systems that aid in increasing the output of manufacturing lines. Powerful machines using machine learning techniques like reinforcement learning have replaced superfluous jobs.

3) ***Transport:*** Transport is a significant area where DS is used. DS is actively making a difference in the transportation industry by creating conditions that are safer for drivers. Additionally, it is essential for improving vehicle performance and giving drivers more autonomy. Additionally, the emergence of self-driving cars has significantly expanded the active use of DS in the transportation sector. DS has established a solid footing in the transportation sector by analysing patterns of fuel usage, driving behaviour, and active vehicle monitoring in great detail. The most popular subjects in the world right now are self-driving automobiles. Vehicle makers may now produce intelligent cars thanks to the introduction of autonomy through reinforcement learning. Additionally, with the use of data science, enterprises may design better logistical routes. Vendors may optimise delivery routes and offer an appropriate distribution of resources by taking into account several factors, such as customer profile, geography, economic indicators, and logistics. Additionally, a number of transportation businesses, like Uber, are embracing data science to optimise prices and give their clients better services. They properly anticipate the price based on variables such as the weather pattern, the availability of transportation, the number of clients, etc., using advanced predictive techniques.

CONCLUSION

The use of Python in the area of data science is the subject of this paper. In today's environment, everyone uses DS, whether knowingly or unknowingly. We also believe that DS, like social science and natural science, will evolve into a new sort of science. In the future, the evolution of DS may be influenced by different technologies.

References

[1] Stančin, I., & Jović, A. (2019, May). An overview and comparison of free Python libraries for data mining and big data analysis. In 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (pp. 977-982). IEEE.

[2] Murtagh, F., & Devlin, K. (2018). The development of data science: implications for education, employment, research, and the data revolution for sustainable development. Big Data and Cognitive Computing, 2(2), 14.

[3] Cao, L. (2017). Data science: a comprehensive overview. ACM Computing Surveys (CSUR), 50(3), 1-42.

[4] Kudande, D. V., & Shah, B. B. A REVIEW PAPER ON FORMATION OF DATA SCIENCE AND FUNDAMENTALS

[5] Siva Jyothi, P. N., & Yamaganti, R. (2019). A review on python for data science, machine learning and IOT. International Journal of Scientific & Engineering Research, 10.

[6] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ...& Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature methods, 17(3), 261-272.

[7] Butwall, M., Ranka, P., & Shah, S. Python in field of data science: A review. International Journal of Computer Applications, 178, 20-24.

[8] Saabith AS, Vinothraj T, Fareez MM. A Review on Python Libraries and Ides for Data Science.

*e*ISSN1303-5150