



An Implementation on Python for Data Science and Machine Learning

Dr. Priyanka Sisodia, Dr. Bhaskar Seth

Geetanjali Institute of Technical Studies, Dabok, Udaipur (Rajasthan).

Abstract-Python is a scriptable and interpreted language for both learning and real-world programming. Python is a powerful high-level language developed by Guido van Rossum. In this paper, we will provide an introduction to the main Python programming software tools used for Data science, Machine learning techniques. Briefly, this paper will first introduce Python as a language, and give introduction about Data science, Machine learning and then describe packages that are popular in the Data science and Machine learning sectors, such as NumPy, SciPy, TensorFlow, Keras, Matplotlib etc. From there, we will move to show the importance of python for building data Science applications. We will use different code examples throughout. To aid the learning experience, execute following examples contained in this paper interactively using Jupiter notebooks.

Keywords: Machine learning · Data Science · Tools · Languages · Python

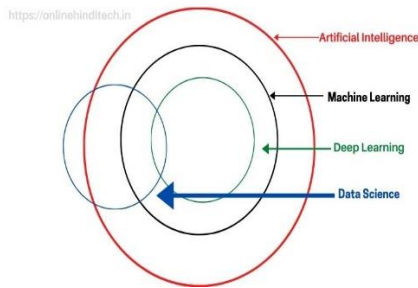
Introduction

1.1 Introduction to python Language

The code in fewer lines that is not possible with other languages. The important feature in Python programming is it supports multiple programming paradigms. Python provides a large collection of comprehensive standard library is extensible. The main features of Python are Simple and easy to learn, High level programming language, Platform independent, Portability, Dynamically typed, Both procedure oriented and Object oriented, Interpreted, Extensible, Embedded, Extensive Library.



In this paper we wish to give brief idea of python in the area of Data science and Machine learning. Python is known to have sufficiency of libraries that assist with data analysis and scientific computing. For example, we can build python application which help data analysts to analyze large amounts of data for scientific computing. The necessities for this paper are basic under-standing of statistics, as well as some



Similarly, companies like Google, YouTube, and Facebook give the video and search you are watching as input to the Machine Learning Model and machine learning is taught by writing a program that whenever a user of this name searches anything, that Showing ads to the user accordingly and recommending videos accordingly.

How Machine Learning Algorithm Works?

Algorithm is used to create and teach Machine Learning Model (ML Model). Algorithm are such steps that ML Model learns by observing and working according to the same algorithm.

Although there are many algorithms to teach machine learning models, but mainly three algorithms are used.

1. **Supervised Learning Algorithm**
2. **Unsupervised Learning Algorithm**
3. **Reinforcement Learning Algorithm**

Using the Supervised Learning Algorithm, some datasets (such as apple color red, weight 20 grams, shape round and height 5 cm) are given to the computer program i.e. ML Model. Using these datasets, computer program (ML Model) Output is given. Two types of dataset are given to ML Model, first is Feature Data and second is Label Data. Using these data, through algorithm, the ML model is taught how to predict the output, let's understand it with an example. Suppose you have to teach an ML Model how to recognize mangoes. So for this you have to use Supervised Learning Algorithm. Now you have to tell that ML Model that how the mango looks like, for this you will give Feature Data to the ML Model, such as the color of the mango is yellow, the shape is round, and the taste is sweet. Along with this, in ML Model, you have to give Label Data as if any input has something like this then its output will be common, common one is Label Data.

Feature Data —> Color Yellow, Round Shape, and Taste Sweet

Label Data —> Common

Now when you have given feature data to the ML Model using Supervised Learning Algorithm, it has been told that if the color of something in an input will appear yellowish, round in shape, and taste sweet, then the prediction of the output will be common . You used Feature Data and Label Data to teach this machine learning system, so this algorithm is called Supervised Learning Algorithm.

Machine learning tools are concerned with endowing programs with the ability to “learn” and adapt. Because machine learning is typically used to process large volumes of data, you may want to choose a powerful low-level language. However, if you’re only



2. Objectives of Study

1. To identify the features of Python Programming.
2. To investigate python modules for Data Science like Numpy which is used for matrix and vector.
3. To focus on python modules for Machine learning like Tensor flow numerical computations for machine learning, Key areas for neural networks and machine learning.

3. Related Works

3.1 Basic Features of Python

Easy to Learn: Python this language is like C, C++ and Java language. Due to not having many keywords in this language, it can be learned quite easily.

Easy to Understand: Due to the syntax like Plain English, it is understood very well.

Easy to Read: Python’s program provides ease of reading without being too complicated.

Simple: Python language is very easy to read and understand. This language is mostly like plain English in reading.

Portable: Python’s program is executed by porting it from one platform (OS) to another.

Large Number of Libraries: Python has many libraries so that no specific code has to be written separately.

1.2 Python for Data Science

Data Science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Data science practitioners apply machine learning algorithms to numbers, text, images, video, audio, and more to produce artificial intelligence (AI) systems to perform tasks that typically require human intelligence. In many scenarios, Python is the programming language of choice for the daily tasks that data scientists deal with, and is one of the top data science tools used across industries. For data scientists who need to incorporate statistical code into production databases or integrate data with web-based applications, Python is often the ideal choice. It is also ideal for implementing algorithms, which is something that data scientists often need to do.

Easy to learn

The most attractive quality of Python is that anyone who wants to learn it - even beginners - can do so quickly and easily and this is one of the reasons why learners prefer Python for data science. It also works well for busy professionals who have limited time to spend learning. When compared to other languages, R, for example, Python boasts of a shorter learning curve with its easy-to-understand syntax.

Scalability

Unlike other programming languages, such as R, Python excels when it comes to scalability. It is also faster than languages like Matlab and Stata. It facilitates scale because it gives data scientists flexibility and multiple ways to tackle different problems—one of the reasons YouTube migrated to the language. You can find Python in many industries, powering the rapid development of applications for all kinds of use cases.

Python Community

One of the reasons Python is so famous is a direct result of its community. As the data science community continues to adopt it, more users are volunteering to build additional data science libraries. It is only driving

the creation of the most modern tools and advanced processing techniques available today, which is why most people are preferring Python for data science.

1.2 Python for Machine learning

Machine learning is a type of learning in which the machine learns on its own without being explicitly programmed. It is a type of application of AI that gives systems the ability to learn and improve on their own from their experience. Here we can create a program which is made by integrating the input and output of the same program.

Conclusion

In this paper we have presented usage of python as a tool in various research areas like DataScience, Machine learning and IOT. Along with Python language, there are many other languages are used for Data science, Machine learning and for developing iot devices like Java, C++ etc. But right now most of the developers use python scripting language than Java, C++. Because of its easy syntax, secure coding, and its simplicity. When it comes to robust and performance, developers choose Python. With respect to the future work there is still huge space for this language to serve other upcoming research areas because of its features like simplicity, extensive library, inbuilt and extensible. In future we will propose python as a powerful tool which is used by many research communities.

Acknowledgement

We would like to thank Department of Computer Application, Geetanjali Institute of Technical studies for encouragement and support in writing this paper.

References

- Jordan, M.I., Mitchell, and T.M.: Machine learning: trends, perspectives, and prospects. Science 349(6245), 255–260 (2015)
- Le Cun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436–444 (2015)
- Holzinger, A., Dehmer, M., Jurisica, I.: Knowledge discovery and interactive data mining in bioinformatics - state-of-the-art, future challenges and research directions. BMC Bioinform. 15(S6), I1 (2014)
- Wolfram, S.: Mathematica: A System for Doing Mathematics by Computer. Addison Wesley Longman Publishing Co., Inc., Boston (1991)
- Engblom, S., Lukarski, D.: Fast MATLAB compatible sparse assembly on multicore computers. Parallel Comput. 56, 1–17 (2016) https://www.researchgate.net/publication/330513589_Internet_of_Things_IOT
- Python Machine Learning: A Guide to Started | Built <https://www.researchgate.net/publication/330513589Data Science>
- <https://www.techaheadcorp.com/blog/top-6-programming-languages-for-Data science-projects/>
- <https://www.google.com/role-of-python-in-iot-development> <https://nbviewer.jupyter.org/github>