Python: The Most Advanced Programming Language for Computer Science Applications

Akshit J. Dhruv, Reema Patel and Nishant Doshi

Computer Science and Engineering, Pandit Deendayal Petroleum University, Gandhinagar, Gujarat

Keywords: Python, Python libraries, Memory allocation, Data Structure, Framework.

Abstract: In last few years, there has been an advancement in programming languages due to different libraries that are introduced. All the developers in this modern era prefer programming language that provides a built-in module/library which can make their work easy. This paper describes the advancement of one such language "Python" and it's increasing popularity through different statistical data and graphs. In this paper, we explore all the built-in libraries for all different computer science domains such as Data Science, Machine Learning, Image Processing, Deep Learning, Natural Language Processing, Data Visualization, Cloud Computing, Speech recognition, etc. We have also included Memory management in Python. Different frameworks for Python which can make the front-end work easier are also mentioned.

1 INTRODUCTION

In 1991, Python language was developed by Guido van Rossum. There is an interesting story behind giving the name "Python" to the programming language. At the time of development of python, the developer was reading the script "Monty's Python Flying which is a BBC series. While reading this book he got an idea to name the programming language as "Python" to have a short and unique name. Python is object- oriented, interpreted, and interactive programming language. It provides high-level data structures such as list, tuples, sets, associative arrays (called dictionaries), dynamic typing and binding, modules, classes, exceptions, automatic memory management, etc. It is also used for parallel computing system and has a comparatively simple and easy syntax for coding and still it is a powerful programming language. Python has the interpreter for java known as JPython, which is similar to the interpreter for C language. Python has many advantages over any other languages, like it has varieties of library which reduces the code to one-third for programmer and due to this Python has reached at the +highest peak in terms of Machine Learning. Difficulty is faced by many while solving problems(Lawan et al, 2015), this research will help providing knowledge about different libraries and motivate them to use Python.

2 DATA STRUCTURE

Data structure means organization, management of data and also it is a storage format which provides efficient access and modification. In general, it contains relation among them, and the functions or operations that can be applied to the data.

- Cython: It helps in improving the speed of the implementation of the code.
- PYTables: It is used in maintaining hierarchical datasets and is also used to maintain an extremely large amount of data.
- Tree Dict: It works as a container for python to simplify the bookkeeping surrounding parameters, variables, and data. It is very stable and fast at work.

Table 1: Different Data Structure in Python.

Type	Definition	Symbol	Example
List	A list is a mutable data structure, ordered sequence of element.	It is defined by square braces [ ].	List=[1,2,3]

Dictionary	Dictionary are also called HashMap or associative arrays, which means that an element of the list is associated with the definition, rather like Map in Java.	It is defined by brackets {}	Dict= {1:"a",2:"b" }
Set	It is a collection of unordered and unique immutable objects.	It is defined by brackets {}	Set={1,2,3 }

3 BUILT-IN LIBRARIES IN PYTHON FOR COMPUTER SCIENCE APPLICATIONS

3.1 Data Science

Data Science is to develop a different approach to record, store, and analyse the data and using this data to get effective information. Data science aims at achieving ideas and knowledge from any type of data.

Python provides number of libraries for the same as listed below:

- Matplotlib: 2D plot graphs can be made using Matplotlib library.
- Pandas: Data analysis in finance, statistics, social science, and engineering require different types of data structure and tools which are provided by Pandas. (https://pypi.org).
- NumPy: It is the basic library for scientific computing in Python. (https://pypi.org) Multidimensional arrays and matrices can be done using objects in NumPy, and also routines are provided which allows developers to compute advanced mathematical and statistical functions on those arrays with code if possible. It is also used in Data Structure.
- SciPy: Manipulation and visualization of data is done using a high-level command provided in SciPy. Functions for solving Integrals numerically, computing differential equations, and optimization are included in the package. The library SciPy is also used in Image processing.
- IPython: Using Ipython, an efficient interactive shell gets added along with the functionality of Python's interpreter that has

the capability of adding rich media, observations, shell syntax, backup of command history, and tab completion. (https://pypi.org) It is also used in debugging by using IPython as fix interpreter. The usage of Mathematica or MATLAB makes it comfortable to work with IPython. It is also used in Data Structure.

- Pygame: Video games are created easily using Pygame. The library has computer graphics and sound libraries which are specially made for python programming language.
- SQLAlchemy: It provides a common interface for creating and executing database-agnostic code without the need of writing SQL statements. It is also used in data structure.
- Scrapy: This library is used to design web scraping, and also it can be used to get data using APIs or it is used as a general-purpose web crawler.
- Pywin32: This library is used to create COM objects and the Pythonwin environment.
- wxPython: GUI toolkit for the Python programming language can be obtained by this library. Applications made using this has native appearance on all platforms.
- Flask: It allows you to build websites and web apps very fast and efficiently.
- Nose: It runs tests or directories whose name includes "test" at the end of the word. To ease out the print- style debugging, it includes captured stdout output from failing tests.
- Sympy: It is used for symbolic mathematics. It tries to keep the code as simple as possible in process of making a full-featured computer algebra system (CAS).
- Fabric: Fabric along which is acting as library for Python, is also a command line interface tool for increasing the use of SSH for the application arrangement or systems administrations. The main use of this library is to create a module which contains one or more functions, and then executing them through fab command-line tool.
- Pillow: Python Imaging Library which adds the support for different options like opening, manipulating data, and saving images as different file formats. It is also used in Image processing.
- Statsmodels: Statistical Models can be estimated using this library. Also it can explore data and perform statistical test. It is also used in machine-learning.

3.2 Machine Learning

Machine learning can also be considered as a subset or part of Artificial Intelligence that can learn automatically and make changes itself from the experience without being externally programming it.(Machine Learning and Deep Learning frameworks and libraries for large-scale data mining).

- Keras: It is a neural networking API and it is to execute for the machine learning beginners to build and design neural networks. It is also used in deep learning.
- Shogun: For a wide range of efficient and unified machine learning methods, Shogun library is used which is an open source library.[5]
- XGBoost: XGBoost is decision tree that uses the algorithm to solve the predictive modelling problems, and this algorithm is efficient and speedy.
- Scikit-learn: It is used for classical ML algorithms. It supports direct and indirect learning algorithms and also be used for data mining and data analysis. It has many applications but majorly unit-testing and self-verification are done using Scikit-learn to detect and diagnose different types of error.
- CatBoost: It will read the documents and analyse the model and data with CatBoost analysing tools.
- PyTorch: PyTorch allows the user to work on Tensors and GPU accelerations by implementing it in C with a wrapper in Lua. It also can create computational graphs.
- Eli5: It is a python library which is used to debug classifiers in machine learning and explain the predictions.
- MlPy: It is the GNU based scientific library and make the extensive use of Cython language.
- Nilearn: Statistical learning on Neurolmaging data can easily be learned and understood using Nilearn library.
- TensorFlow: It is used for high-performance numerical computation. It can develop many artificial Intelligence applications by implementing the deep neural networks. It is also used in Deep Learning.(Machine Learning and Deep Learning frameworks and libraries for large-scale data mining)

According to the GitHub survey from January 2018 to December 2018, Python is at the top in terms of machine learning compared to C++, Java, JavaScript, etc. (https://github.com).

3.3 Deep Learning

Deep Learning can also be called part of Machine learning. It has a layer of Artificial Neural Network which can learn the unstructured or unlabelled data. (Machine Learning and Deep Learning frameworks and libraries for large-scale data mining).

- Apache MxNet: It permits to use mix symbolic and crucial coding to increase productivity and efficiency Inside MxNet there is a modern dependency scheduler that will help to automatically parallelizes both symbolic and imperative operations quickly.
- Caffe: Expression, modularity, and speed are the key features of this library.
- Fastai: It simplifies the training of neural nets very quickly and with accuracy and using the latest technique. It includes the support for text, vision, and tabular models. (Machine Learning and Deep Learning frameworks and libraries for large-scale data mining).
- CNTK: Neural networks are defined in a directed graph by a series of computational steps that are described by the toolkit of CNTK. In this graph, input values are represented by leaf nodes or network parameters and other represent matrix operation which took place on the inputs.
- TFLearn: High level API is provided to TensorFlow using this library. This library helps to get the quick outputs of the experimentation and also the process is transparent.
- Lasagne: It is used to build and train neural networks using Theano. Convolutional Neural Networks (CNNs), recurrent networks are supported by this library.
- Elephas: Deep learning's distributed models can run at
  scale with the use of spark using this library.
- Theano: Mathematical expression for multi-dimensional arrays can be optimized, evaluate and can also be defined in this library. Theano is also used in Machine Learning.

3.4 Image Processing

Image processing is specially used to do some operations on an image to get a better-quality image or to find some useful information from it. It works like signal processing in which we take input as image and output may vary, like it can be image or characteristic features which is associated with that image.

- Scikit Image: It is a collection of algorithms for image processing and uses NumPy arrays as image objects. It includes algorithms for geometric transformation, segmentation, colour space, analysis, manipulation, filtering, morphology, feature detection, etc.
- Open cv-python: It is developed by Intel for real-time image & video analysis and processing.
- Mahotas: Functions such as morphological operations, modern computer vision functions and filtering for the advanced computation and includes the interest point detection also.
- Cairo: It acts as a 2D graphics library for python and also supports many output devices. Using display hardware acceleration, it gives continuous output on all connected devices.

3.5 Game Development

Game Development is used to create games and describes the design, development, and release of a game. Before game development, it is important to think about the game mechanics, rewards, player engagement, and level designing.

- Pyglet: It is a cross-platform windowing and multimedia library for Python, developed to create games and other visually rich applications. It has feature which can load images, sound, music, and video in almost any format.
- Arcade: It is used to create 2D video games. It helps the developer to create the 2D games without learning complex frameworks.
- Rabbyt: It is a sprite library for python which provides fast performance with an easy to use but flexible API.
- Pymunk: It is a pythonic 2D physics library that can be used whenever there is a need for 2D rigid body physics from Python.
- Pybox 2D: It is purely 2d engine written primarily for games. It includes features like circles, up to 16 sided polygons, thin line segments, controllers, basic breakable bodies, and pickling support.
- Panda 3D: It is used for 3D rendering and game development. It also support automatic shader generation, which means that we can use normal maps HDR, gloss maps, cartoon shading glow maps.

3.6 Networking

Python provides two-level access to networking. One low level, in which one can access the basic socket support in the same OS that permits implementation for clients and servers to do connection-orientation and connectionless protocols.

- asyncio: It provides base for writing the existing code in a single sequence using coroutines, multiplexing I/O access over sockets and different resources, which are running network clients, servers, and other related primitives. To detect common issues during development, debug mode is enabled.
- Tftpy: It includes client and server classes and create a TFTP server/client to receive/send files.
- Telnet lib: It provides a telnet client implementation, so it represents a connection to a Telnet Server.
- Parmiko: It is an implementation of SSHv2 protocol, providing the functionality of client and server both. SFTP client and server mode are supported.
- Requests: All kinds of HTTP requests are sent in python using this module.

3.7 Natural Language Processing

Natural Language Processing shows the connection between human language and computers. It is used in businesses and it is a very important term in every engineer's life.

- Gensim: It is used for topic modeling, similarity revival and document indexing with large corpora. It includes features like all the algorithms are memory- independent compared to the intuitive interfaces, corpus size, efficient multicore implementations of popular algorithms, distributed computing etc.
- Textblog: This library is used for processing written data. It provides API for part-of-speech tagging, sentiment analysis, noun phrase extraction, classification, translation, WordNet integration, word inflection, parsing, add new models or languages through extensions.
- SpaCy: It is a Natural language Processing library of Python which contains pre-trained statistical models, word vector and also it has support tokenization for 49+ languages.
- Vocabulary: It is a Python library, which is used to get the meaning, synonyms, opposites, part of speech, translations for a given word.

- PyNLPl: PyNLPI is also read as "Pineapple", which is used in extracting a sequence of n items in a list and also used in building a simple language model. This library covers a large area in the field of working with FoLiA XML.
- NLTK: Lexical resources such as WordNet requires an interface to run in Python which is provided by NLTK and also provides library called text processing for tokenization, classification and stemming. It has built-in function which provides practical ideas to programming for language processing.
- CoreNLP: Using this library it is easy to apply groups of linguistic analysis tools to a piece of text. It includes many tools such as: Part-of-speech recognizer (POS) tagger, the conference resolution system, the parser, the named entity recognizer, sentiment analysis, the open information extraction tools, and bootstrapped pattern learning.
- Pattern: It has tools for data mining, natural language processing, machine learning, network analysis using graphs centrality and visualization.

3.8 Cloud Computing

Cloud Computing uses a network of remote server which is hosted on the Internet to store, process, and manage the data, rather than a local server or a personal computer.

- Apache Libcloud: It is a python library for cloud computing and it does work of hiding difference between different cloud provider APIs and allows us to manage different cloud resources through a unified and easy to use API.
- google-api-python client: It allows us to work with Google APIs such as Google+, YouTube or Drive on our server. They are officially supported by Google.

3.9 Data Visualization

Data visualization is used to represent the information in the form of a chart, diagram, and pictures.

- Seaborn: It is a matplotlib based library that provides an interface of high-level for making catchy and informative statistical graphics.
- Bokeh: It is used as an interactive visualization library that can target web browsers for representations. We can pass all

types of data such as Python lists, tuples, NumPy arrays or Pandas Data Frames to make the plots.
- Py.gal: Formats for the vector graphic are SVG whose charts can be made from python using a library called PyGal.
- Ggplot: It is a very helpful library for the programmers who are coming from the R background and used ggplot2 in it, as the same Ggplot is used in python.
- Plotly: It is an interactive library use for plotting which supports 40 different types of graphs covering a wide range of financial, statistical, geographic, scientific, and 3 dimensional.
- Missingno: Many times the datasets are missing and they are represented by NaN (not a number). So this library provides a way to visualize the distribution of NaN values. It is compatible with Pandas.
- Leather: It makes the work done fast but not with perfection, like someone needs charts but doesn't expect any type of accuracy or perfection can use this library.
- Pydot: It provides a complete interface to create, handle modify and process graphs.

3.10 Speech Recognition.

Speech recognition is used to convert the human voice to computer understandable language using different software or different hardware. It has many applications, like to give command to computer do perform any particular task without even writing or working physically.

- CMU sphinx: It contains the best toolkit with many different tools used to build speech applications.
- Google Speech Recognition: Provides facility likes
  - o Configure Microphone.
  - o Set chunk size.
  - o Set the sampling rate.
  - o Set device id to selected microphone.
  - o Allow adjusting for unknown noises.
- Microsoft Bing Voice Recognition:
  - o Microsoft Heera
  - o Microsoft Zira
  - o Microsoft David
  - o Microsoft Mark
  - o Microsoft Ravi

  are some of the in-built voice used in speech recognition app

- Houndify API: It is used in creating a client that converts speech to text in under a few minutes. It provides a simple way for developers to use the platform for its speech to text capabilities through the speech to text only domain.
- Api.py: It is used to combine speech recognition with natural language processing.
- Ppytsx3: It is a text to speech python library which works without internet connections or any type of delay.
- PyAudio: It is the cross-platform audio I/O library. This library helps to play and record audio streams on several platforms.

3.11 Cryptography

This library provides a set of procedure to decrypt messages and to secure communications among computer system, matplotlib, etc.

- PyNacl: It is a bounded library called libsodium which is a collection of the Networking and cryptography library.
- PocketProtector: It is a library that contains a secret management system.
- Pycryptodome: It is low-level security providing library but still used for home purposes.

4 MEMORY MANAGEMENT IN PYTHON

Memory management is an important factor that everyone checks before choosing any programming language. Memory management is allocating particular block to programs and reduce the overall space required and increases system performance.

Comparison in memory allocation in other languages and Python: int a = 10; int b= 10;Then in C language, it is stored as variable and both a and b will be provided a different memory space.

While in Python it is stored as reference. When we enter value as 10 the reference count of 10 becomes 1 and when the value of b is added as 10 and the reference count becomes 2.

Table 2: Difference between Python and C.

| Python	C Language
No data is stored during compile time.	Data is stored in "Stack" during compile time.
During run time the data is stored in "heap".	During run time the data is stored in "heap".
There is no use of variable in storage.	Variable is also stored in memory.
All the values are stored as reference making the memory optimized.	All the values are stored using memory location as well as variable in which it is stored.
After the execution of The program, if the Reference count is 0 then The garbage collector will release the memory on its own.	After the execution also the memory area stays occupied unless the command is not given externally to free the memory.

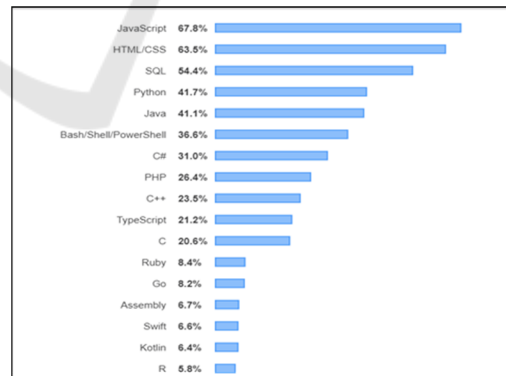### 4.1 Second Section
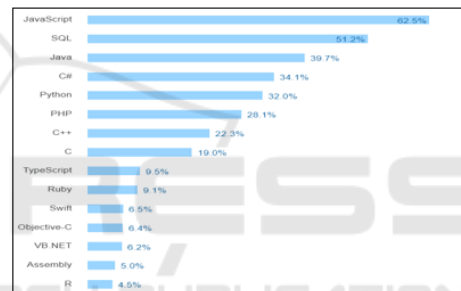
Figure 1: Survey of 2017.

Figure 2: Survey of 2019

Comparing the graph of 2017 and 2019 we can see it has raised its position from 5th rank to 4th rank. Also, the number of users of Python in 2017 was 32% which increased to 41.7% in 2019.
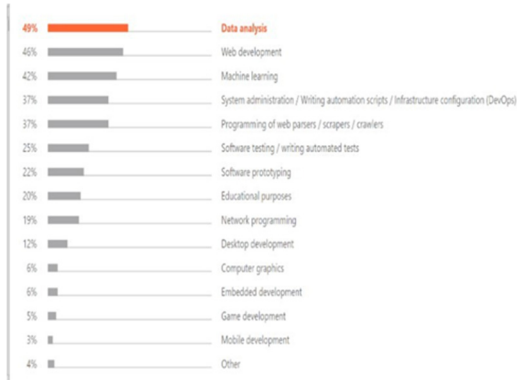
4.2 Usage of Python

Figure 3: Graph showing usage of Python

4.3 Average Python Programmers Salary by States: 2019

Figure 4: Salary chart by states.

Here we can see that California has the highest paid salary to the programmer and slowly there will be others joining them.(https://www.daxx.com/blog/development-trends/python-developer-salary-usa)
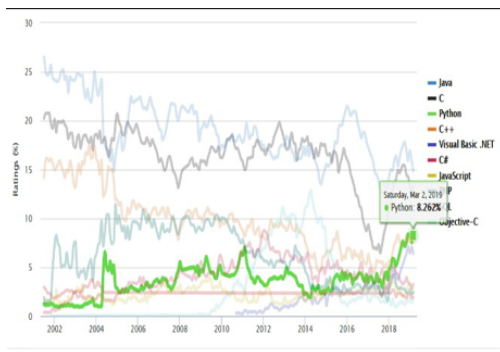
4.4 Python's Rating

Figure 5: Rating of Python from time it was developed.

This graph shows the increment in rating since the time it is developed.

4.5 Top Websites Built using Python

• The top websites built are Instagram, Spotify, Netflix, Google, Uber, Dropbox, Pinterest, Instacart, Reddit, and Lyft.(https://learn.onemonth.com/10-famous-websites-built-using-python/).

5 THIS SECTION INCLUDES INFORMATION ABOUT THE DIFFERENT FRAMEWORKS USED IN PYTHON

Frameworks: A collection of different modules/packages which are used by developer to write web-applications or services without requirement of handling minor details such as protocols, socket, or process management. (Kumar, Dahiya)

5.1 Full Stack Framework

It is a framework that tries to provide nearly everything i.e. from web serving to database management right down to HTML generation — which a developer could need to build an application.(Nithya et al).

Few Full Stack Frameworks :

- *Django*   - *Giotto*   - *Zopez*
- *Turbo Gears*   - *Grok*   - *kiss. Py*
- *web2py*   - *Pylon*   - *Lino*
- *Cubic web*   - *Reahi*   - *Nagare*
- *Tornado*   - *wheezy.web*   - *Pylatte*

5.2 Non-full Stack

These frameworks do not provide extra functionalities and features to the developers. They have to add huge code and components manually here.

Few Non-Full Frameworks are:

- *Bottle*   - *AppWsgi*   - *Clastic*
- *Cherry.Py*   - *BlueBream*   - *Divmod*
- *Flask*   - *More Path*   *Nevow*
- *Hug*   - *Bobo*   - *Falcon*
- *Pyramid*   - *Bocadillo*   - *Growler*

6 CONCLUSION

Python is growing rapidly and has reached to 3rd rank in terms of best programming language. It is seen that; Python has many libraries which makes it unique from other programming languages. It's popularity and ratings are increasing day by day along with the demand of Python programmers all over the world.

REFERENCES

Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey.

Lawan, A. A., Abdi, A. S., Abuhassan, A. A., Khalid, M. S., 2019. What is Difficult in Learning Programming Language Based on Problem- Solving Skills?," International Conference on Advanced Science and Engineering (ICOASE), Zakho - Duhok, Iraq.

Lo, C., Wu, C., 2015. *Which Programming Language Should Students Learn First? A Comparison of Java and Python," 2015 International Conference on Learning and Teaching in Computing and Engineering, Taipei, pp. 225-226.*

Prof. B Nithya Ramesh, Aashay R Amballi, Vivekananda Mahanta, "Django Python Framework", International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online) Vol. 6, Issue 2.

Krishan Kumar, Sonal Dahiya, "Programming Language a Survey, International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 5.