

COMPARISON BETWEEN PYTHON, JAVA AND R PROGRAMMING LANGUAGE IN MACHINE LEARNING

Ben Johnson*1, Dr. Anjana S Chandran*2

*1Student, Department Of MCA, SCMS School of Technology And Management, Ernakulam, Kerala, India.

*2Associate Professor, Department Of MCA, SCMS School of Technology And Management, Ernakulam, Kerala, India.

ABSTRACT

Comparison of programming languages may be a common topic of dialogue among software engineers. Multiple programming languages are designed, specified, and implemented per annum so as to stay up with the changing programming paradigms, hardware evolution, etc. during this paper we present a comparative study between three programming languages used for Machine Learning. Machine learning is that the study of computer algorithm that improve automatically through experience and by the employment of information. it's seen as a component of computing. Machine learning algorithms build a model supported sample data, noted as "training data ", so on form predictions or decisions without being explicitly programmed. The complexity of Machine Learning (ML) models and also the frameworks people are using to create them has exploded together with ML itself. Considering the accelerating rate at which software machine learning is progressing, emerging technologies like Java, Python , and R. These languages are compared under the several characteristics.

Keywords: Python, Java, R, Machine Learning.

I. INTRODUCTION

Machine Learning has started playing employment in our daily lives. Machine learning is a subset of AI (AI) that gives systems the ability to automatically learn, think and improve from experience without being explicitly programmed by a human. Machine learning focuses on the event of computer programs that can access data and use it to be told for themselves. The first aim is to permit the computers learn habitually without human interface or assistance and adjust actions accordingly.

Thousand different languages are created within the previous number of years. Some languages enjoy wide popularity and also introduce new features. Each language has its advantages and drawbacks. this work provides a comparison of various properties, paradigms, and features utilized by a pair of popular programming languages: Java, Python and R With these style of languages and their widespread use, software designer and programmers should in contact in mind of the benefits and disadvantages each language could rouse their software solution and use caution once they create rational decisions. These languages are compared under several characteristics. Other criteria rather similar to the programming effort, run time efficiency, memory consumption, and database connectivity are disclosed by implementing and running the identical set of programs using all the languages under study

In three languages, Python, R, and Java are dynamically typed, have a command interface for the interpreter, and are available with great number of libraries to support scientific and technical computing. Conveniently, these languages also offer great solutions for simple plotting and visualizations. Combined with interactive notebook interfaces or dynamic report generation engines data analysis and documentation has never been easier.

II. OVERVIEW

2.1 Python

Python could even be a suitable language for both learning and global programming. Python would be a strong high-level, object-oriented language created by Guido van Rossum.. Python is termed as a very user-friendly and beginner-friendly language in the recent times. Python has gained popularity for being a beginner-friendly language, and it has replaced Java as the most popular introductory language. Python could even be a general-purpose, problem-oriented language which is widely employed within the recent times. Python may be a suitable language for both learning and global programming. Its design philosophy emphasizes readability, and

its syntax allows programmers to specific concepts in fewer lines of code. The language constructs enable the user to put in writing clear programs on both a little and enormous scale. Python may be a suitable language for both learning and global programming. Python is an Object oriented scripting language, The run time speed of this language is 71.90secs and the memory utilized is 2.80mb/sec. Python is employed by larger firms principally which will appraise that can evaluate huge data sets. Python had heavily borrowed the module system, exception model and also keyword arguments from Modula-3 language. During its design, the creators had conceptualized the language as being a very extensible language, and hence that they had designed the language to own a little core library which was extended by an enormous standard library. Thus as a result, python is employed as a scripting language because it is easily embedded into any application, though it is wont to develop a full-fledged application. Python supports multi-paradigms like Object-Oriented, Imperative, Functional, Procedural, and Reflective. In Object-Oriented Paradigm, Python supports most of the OOPs concepts like Inheritance, Polymorphism but its lack of support for encapsulation. Python compiles to an intermediary code and this in turn is interpreted by the Python Runtime Environment to the Native machine language. The reference implementation doesn't include a JIT compiler due to which the fastness is slow compared to native programming languages. The Python Runtime Environment manage all the allocation and deal location of memory through the Garbage Collector. When a new object is created, the GC allocates the required memory, and once the thing goes out of its scope, the GC doesn't release memory immediately but instead it becomes eligible for trash collection, which might eventually release the memory. Python may be a strongly dynamic typed language and also supports optional static typing. Python is predominantly used as a scripting language used in developing standalone applications that are being developed with Static-Typed languages, due to the flexibleness it provides thanks to its dynamic typed nature. Python favours rapid application development, which qualifies it to be used for prototyping. To a specific extent, Python is additionally utilized in developing websites. thanks to its dynamic typing and of the presence of a Virtual Machine, there is a considerable overhead which translates to way less performance when we compare with other programming languages .Python Supports dynamic type system and automatic memory management and contains a large and comprehensive standard library. Python interpreters are available for several operating systems.

2.2 Java

Java language was originally developed by James Gosling, Mike Sheridan, and Patrick Naughton within the year June 1991 at Sun Microsystems Sun Microsystems which was initiated by James Gosling and released in 1995 as core component of Sun Microsystems. Java is that the invisible force behind many of the devices and applications used on everyday and power everyday lives. Java could be a wonderful OOPs language. It offers encapsulation through access modifiers like private, public, protected etc. which improves the reliability and security aspect of the code. But compared with other programming languages like C++ and Python, Java doesn't support multiple-inheritance of classes though it supports multiple inheritance through Interfaces. Java is used for mobile-based applications, enterprise-level purpose, for making desktop applications, and for establishing Android apps. Though it had been heavily influenced by C and C++, it had been different in its own ways. It eliminated pointers as its designers thought that developers are using pointers the way it had been not meant to be. Also, it dropped the headache of managing memory as it brought in automatic memory management with the introduction of trash collection. Java also supports Procedural Programming. But java does not let us have global variables or methods outside the class, which could be attributed to one of its design goal. Thus, Java is more of a OOPs language than anything else. Java supports Reflective paradigm with API requests to access, create, modify and add class members. Java supports creation of a unlimited multitasking applications because it consist of limiteless API that supports creating, managing, communicating between threads and processes. It also support many data structures that support atomic access and it also provides many utilities required for thread/processes synchronization .The java code is compiled to bytecode by the java compiler and it is executed on the JVM. As a result, the execution speed was tremendously slow when compared with other languages like C++ . But with the introduction of JIT Compiler, the execution speed was greatly improved. JVM takes care of all the resources that is needed by the program. All the allocation and deallocation of memory is handled by the JVM through the Garbage Collector. Java is a strongly static typed language and also supports dynamic typing thanks to the fact it supports polymorphism and reflection

programming paradigms The advantages of Java being a static typed programming language, is that it allows programmers to find out the errors at the compile time.

Since Java could be a static typed language, most of the coding errors are detected during the compile time because of which the unit testing time is greatly reduced and therefore the user has to test only the business logic Java also gives us the liberty in choosing from the vast library of quality frameworks available for any task that needs to be implemented using java. the foremost important weakness is that's that since it's not a natively executed language, its performance may be a bit less compared with the opposite natively executed languages like C, C++. Not only is it possible to use Java for machine learning and data science application development, but it's also the well-liked option by many developers for variety of reasons. Developers consider the Java Virtual Machine collectively of the most effective platforms for machine learning and data science, because it enables the developer to jot down code that's identical across multiple platforms. It also allows them to make custom tools at a faster pace and features a load of IDEs that help to enhance overall productivity levels. Java makes it an excellent choice for the building of larger or more complex computing and Machine Learning applications, especially once they are being built from scratch. Production codebases are commonly written in Java. Knowing Java helps developers work out how data is being generated, submit merge requests to production codebases, and deploy Machine Learning solutions to production. Java has many libraries and tools available for Data Science and Machine Learning. for instance, Weka 3 could be a fully Java-based workbench popularly used for algorithms in machine learning, data processing, data analysis, and predictive modeling.

2.3 R

R is one in every of the foremost popular and widely-used software systems for statistics, processing, and machine learning. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently maintained by the R Development Core Team. R is out there under the GNU General Public License, and pre-compiled binary versions are provided for various operating systems like Linux, Windows and Mac. it's mainly used for statistical computing. It offers an expansion of algorithms which are heavily utilized in machine learning domain like statistic analysis, classification, clustering, linear modeling, etc. R is incredibly the identical to the S language. R does not allow to do branching and looping but also allow to do modular programming using functions. It allows integration with the procedures written in the C, C++,Net, Python, and FORTRAN languages to improve efficiency. R programming language having a very powerful set of built-in functions, and the ability to write functions which can use expressions as input. The use of R is based on the concept of defining data structures that fit within an object-oriented programming language. Thus, R objects, which can be data structures or functions, have specified classes and methods. This permits advanced programming in which actions within a function are directed by the data type that is being processed. This sophistication permits elegant ways of solving computational problems. It can also lead to many unexpected side effects. Users and programmers in R need to learn how to work with objects and avoid unpleasant effects. R is capable of computing mathematical and statistical management functions. Functions operate data structures, which include data types like lists, vectors, characters, factors, matrices and data frames. The programming style in R is elementary or advanced, invoking many of the object-oriented aspects of the programming environment. The R language may be a fully functional programming environment and users can use built-in functions or write their own functions. variety of other built-in functions assist in visualizing the results of statistical analysis through a comprehensive group of graphing functions. Statistical modeling functions include linear models, analysis of variance, generalized linear models and nonlinear statistical procedure and maximum likelihood models.

III. PERFORMANCE RESULTS

3.1 Run Time Speed

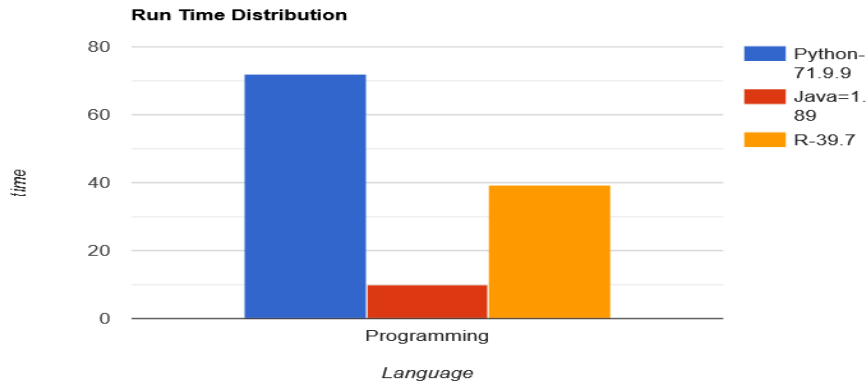


Figure 1: Run Time Speed

Some languages can be compiled fast, and others are relatively slow. Without looking closely at the language and how the compiler works there isn't a simple way to predict if a language will be slow to compile, but there are two major considerations. *Structure* - The structure of the language itself impacts compile time. *Dynamic vs Static* - How much is done at compile time makes a difference, though often not a large difference. Languages like C++ do most of all name and type resolution at compile time, which may increase it's time, whereas Java does some at runtime, and dynamic languages like JavaScript do most things at runtime. In the following Python programming Language Run time speed for faster comparing with the other languages.

3.2 Memory Utilization

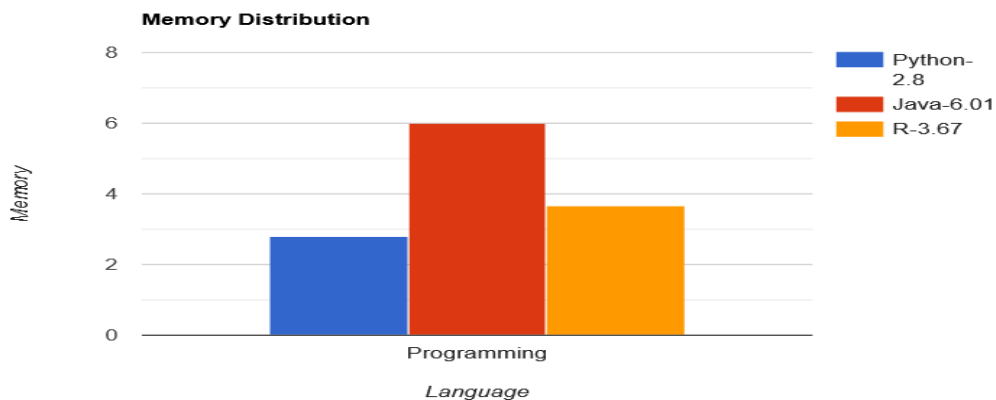


Figure 2: Memory Utilization

Memory management is the process of controlling and coordinating the way a software application access computer memory .The process of allocating, deal locating, and managing memory effectively is called memory management. It basically refers to the allocation of memory when required by the program and then frees it when it's no longer required. There are two related tasks of memory management:

- 1) Allocation: Allocation of memory on the request of the program for the block of memory. This block of memory is received from the operating system. The part of the memory manager that performs this, is known as the allocator.
- 2) Recycling: Once the memory blocks have been allocated, but no longer required, then they are recycled either by manual memory management in which the programmer decides this or by automatic memory management in which the memory manager does this work .In the Following Java Has the best memory utiization when comparing to other two languages.

IV. TABLE OF COMPARISON

Table 1. Comparison of Programming Language

Aspect	Python	R	Java
First Release	1991	1995	1991
Authors	Guido Van Rossum	Ross Ihaka and Robert Gentleman	James Gosling, Mike Sheridan
Stable Version	3.9.5	3.6.2	16.0.1
Number Of Packages	199816	15102	85367
Main Implementation Language	C	C and Fortran	C
Run Time Speed	71.9	1.89	39.7
Primitive Datatype	Integer,Float,Strings, Boolean	Numeric,Int,Character, Logical	Int,byte,short, long,float,double, boolean and clear
Object Oriented	Yes	Yes	Yes
Code Structure	Based on Indentation	Free Style	Free Style
Bounds checking	Run Time	Run Time	Run Time
Memory Utilization	2.8	3.67	6.01
Programming paradigms	Multi-pradigm: object-oriented, imperative ,functional, procedural, reflective functional	Array,object-oriented,imperative, functional,procedural, reflective	multi-paradigm: object-oriented, structured,imperative, generic,reflective,
Exception Handling	Yes,via except/finally block we can have multiple except block	Yes try,trycatch,with trycatch	Yes,via except/finally block we can have multiple except block

V. CONCLUSION

In this comparative study, several experiments have been made and a number of specific criteria were considered to compare our selected programming languages. This study has revealed that, due to their internal design and structure, each language is best suited for a specific application domain. The technology world is expanding extremely with each passing year and months .The various programming models can be approached indiuadually of one another from modular or object-oriented program structure to low-level systems programming.. Python language is faster than both Java and R. The run time speed of the Python is relatively more compared to Java and other languages. Java has demonstrated better for web programming and J2EE as well as desktop applications. Python can be used as a scripting language with other static typed programming languages to develop enterprise application. It can also be used for rapid prototyping as, with python we can achieve less code to task ratio. R is used a lot in academia, so people with that background or doing academic

research usually pick R. Finally by considering the memory utilization Python is better when compared to other languages the memory used is less.

VI. REFERENCES

- [1] "Programming Language Trends - O'Reilly Radar". Radar.oreilly.com. 2 August 2006.
- [2] Python- The Fastest Growing Programming Language" K. R. Srinath.
- [3] D. Smith, "CRAN now has 10,000 R packages. Here's how to find the ones you need," 27 January 2017. [Online]. Available: <http://blog.revolutionanalytics.com/2017/01/cran10000.html>. [Accessed 31 October 2017].
- [4] Data structure based performance evaluation of emerging technologies — A comparison of Scala, Ruby, Groovy, and Python, Miiti S Bhat, Deepthi G Nair
- [5] Guido van Rossum. (1995), Python Tutorial, Amsterdam, The Netherlands: CWI (Centre for Mathematics and Computer Science).
- [6] Java-ML: A Machine Learning Library Thomas Abeel Yves Van de Peer
- [7] Tim Lindholm, Frank Yellin, Gilad Bracha, Alex Buckley. (February 25, 2013), the Java Virtual Machine Specification, Java SE 7 Edition (Java Series): Addison-Wesley Professional.
- [8] Kendal, Simon, 2008. Object Oriented Programming using Java. ISBN: 978 -87-7681-501-1
- [9] An empirical comparison of seven programming languages, L. Prechelt
- [10] R: A Language for Data Analysis and Graphics, Ross Ihaka, Robert Gentleman
- [11] Evaluating the Design of the R Language, Floréal Morandat, Brandon Hill, Leo Oswald Jan Vitek
- [12] S Machine Learning Made Easy: A Review of *Scikit-learn* Package in Python Programming Language
- [13] Dlib-ml: A Machine Learning Toolkit, Davis E. King
- [14] Python Machine Learning: Machine Learning and Deep Learning With Python, Scikit-Learn, and TensorFlow 2, Third Edition
- [15] <https://jaxenter.com/energy-efficient-programming-languages-137264.html>