PRACTICAL FEASIBILITY OF ANALYTICAL TOOL

Ease of Use:

R is not easy to learn and understand because it takes more time to understand the syntax and handle the error. If one has no or less experience in programming will find it very difficult to learn and implement. Whereas, Python is a very simple programming language with an easy syntax that anyone can learn, understand and implement easily, even those who have just begin programming can grasp it easily and learn quickly [12].

Available Library:

A library is a collection of functions that can be included in the program. It is very essential just like any other function. Python has extensive libraries for every task that ease the development process and saves time. However, R contains more library that is useful for data science and statistics, Python is mainly focused on machine learning, deep learning, and artificial intelligence. There are many power libraries for both Python and R, these libraries update regularly [13].

In this study, To prepare and process the dataset in Python, different libraries like Pandas, Numpy has been used. Whereas, tidyverse and Dplyr are used in the R environment. For dataset exploration, visualization, and analysis matplotlib, seaborn libraries are used in python, On the contrary, the ggplot2 library has been used in R to explore and visualize the dataset. To apply a machine learning model in python scikitlearn library used in python, whereas, caret and Randomforest libraries have been imported in the R environment [14].

You can visualize the code screenshot in Appendix Figure 2 and 3.

Visualization Capability:

Data visualizations are one of the fundamental parts of data analysis. It describes complex information in a manner that everyone can easily identify patterns and correlations. R has many advanced and powerful packages like ggplot2, Lattice for data visualization. These packages make complex raw data set understandable, informative, and eye-pleasing. Whereas, Python has also a large number of interactive packages like Matplotlib, seaborn, pydot, bokeh, etc. for visualizing the data. However, choosing the best and most relevant package can be a very intricate task for the data scientist. R provides better data visualization with less complexity in comparison to python [14].

The collected dataset has been visualized through both python and R. First, look at the data frames generated from the dataset by Python and R. Figure 1 and 4. As we can see that python and R has a different way to visualize the data for exploratory data analysis. If we look at the relationship graph figures 4, 5, 6,7, 8 with the target variable in both python and R.

Computational speed:

Both Python and R are used for data analytics but R is slower than python. You can analyze a large dataset through python with less code and less time. R was mainly created for the statistical task; it has many packages to improve performance. R needs more code to do a single task that reduces computational speed when handling a large dataset [15].

Community support:

R has big community support with more than two million users. There are thousands of developers and programmers around the world who are part of this R community. The programmer and member of stack overflow are also contributing to growing R. You can get more statistical solution on the R community. Popular community support for R such as R-studio community, R-statistic community, stack overflow community [16].

However, Python has also rich community support like R, it has comprehensive online community support without customer service support. The biggest advantage of python, there are millions of developers who can support you when you run into trouble or facing any code error. Python community provides strong support for machine learning, Artificial Intelligence. Popular python communities like pycon, pyslackers, StackOverflow, and so on available to support your work [17].

Statistical correctness:

R provides better support and libraries for statistics as it is specially developed for data statistics. Whereas, Python is best for deep learning, machine learning, and application development and deployment. So it can be assumed that R has more statistical correctness than python because the libraries of R implement a wide range of statistical and graphical techniques for data analysis [18]. However, predicting heart disease applying logistic regression in python provides more accuracy than R. After doing the confusion matrix the accuracy score of the Python logistic regression model is 0.9230, whereas R logistic regression model accuracy is 0.7444. Please see figures 9 and 10

Cost:

Both Python and R is an open-source programming language, most of the times the IDE and the environment used for python and R are comes with free of cost. However, there is some paid version of IDE can add additional cost while developing with Python and R. Pycharm is one of such IDE that offers both community and professional version. The community version comes free of cost whereas the professional version costs. On the contrary, R studio is a popular IDE for development in R, it also offers a paid and free version [18].

System Integration:

Python can be integrated with other languages it supports R functionality via the RPy2 package. For easy development, the python programs are integrated with web apps. On the other hand, R runs programs locally and integration can be very difficult. As R has complex syntax, particularly for the new users it is quite impossible to integrate with another language [19].

Consistency:

To add more functionality to the program, R heavily depends on third party code for many packages. It often creates an inconsistency within the available algorithms. Whereas Python is very flexible and versatile in this respect. The method of producing large packages with fewer features means that the code is consistent across the platform. As we know python provides great community support [18].