UNIT 5   R Programming

Advantages of R Programming

1. Open Source

R is an open-source programming language. This means that anyone can work with R without any need for a license or a fee. Furthermore, you can contribute towards the development of R by *customizing its packages, developing new ones and resolving issues.*

2. Exemplary Support for Data Wrangling

R provides exemplary support for data wrangling. The packages like *dplyr, readr* are capable of transforming messy data into a structured form.

3. The Array of Packages

R has a vast array of packages. With over 10,000 packages in the CRAN repository, the number is constantly growing. These packages appeal to all the areas of industry.

4. Quality Plotting and Graphing

R facilitates quality plotting and graphing. The popular libraries like *ggplot2* and *plotly* advocate for aesthetic and visually appealing graphs that set R apart from other programming languages.

5. Platform Independent

R is a platform-independent language. It is a cross-platform programming language, meaning that it can be run quite easily on Windows, Linux, and Mac.

6. Machine Learning Operations

R provides various facilities for carrying out machine learning operations like *classification, regression and also provides features for developing artificial neural networks.*

7. Statistics

R is prominently known as the lingua franca of statistics. This is the main reason as to why R is dominant among other programming languages for developing statistical tools.

8. Continuously Growing

R is a constantly evolving programming language. It is a state of the art that provides updates whenever any new feature is added.

Disadvantages of R Programming

1. Weak Origin

R shares its origin with a much older programming language **"S"**. This means that it does not have support for dynamic or 3D graphics.

2. Data Handling

In R, the physical memory stores the objects. This is in contrast with other languages like Python. Furthermore, R utilizes more memory as compared with Python. Also, R requires the entire data in one single place, that is, in the memory. Therefore, it is not an ideal option when dealing with Big Data.

3. Basic Security

R lacks basic security. This feature is an essential part of most programming languages like Python. Because of this, there are several restrictions with [R](#) as it cannot be embedded into a web-application.

4. Complicated Language

R is not an easy language to learn. It has a steep learning curve. Due to this, people who do not have prior programming experience may find it difficult to learn R.

5. Lesser Speed

R packages and the R programming language is much slower than other languages like *MATLAB* and *Python*.

The algorithms in R are spread across different packages. Programmers without prior knowledge of packages may find it difficult to implement algorithms.

Applications of R Programming

Some of the important applications of R Programming Language in the domain of Data Science are:

1. Finance

Data Science is most widely used in the financial industries.

R is the most popular tool for this role. This is because R provides an advanced statistical suite that is able to carry out all the necessary financial tasks.

With the help of R, financial institutions are able to perform downside risk measurement, adjust risk

performance and utilize visualizations like *candlestick charts, density plots, drawdown plots,* etc.

R also provides tools for moving averages, autoregression and time-series analysis which forms the crux of financial applications. R is being widely used for credit risk analysis at ANZ and portfolio management.

Finance industries are also leveraging the time-series statistical processes of R, to model the movement of their stock-market and predict the prices of shares. R also provides facilities for financial data mining through its packages like *quantmod, pdfetch, TFX, pwt,* etc. R makes it easy for you to extract data from online assets. With the help of *RShiny*, you can also demonstrate your financial products through vivid and engaging visualizations.

2. Banking

Just like financial institutions, *banking industries make use of R f*or credit *risk modeling and other forms of risk analytics.*

Banks make heavy usage of Mortgage Haircut Model that allows it to take over the property in case of loan defaults. Mortgage Haircut Modelling involves sales price distribution, the volatility of the sales price and the calculation of expected shortfall. For these purposes, R is often used alongside proprietary tools like SAS.

R is also used in conjunction with Hadoop to facilitate the analysis of customer quality, customer segmentation, and retention.

Bank of America makes use of R for financial reporting. With the help of R, the data scientists at BOA are able to analyze financial losses and make use of R's visualization tools.

3. Healthcare

Genetics, Bioinformatics, Drug Discovery, Epidemiology are some of the fields in healthcare that make heavy usage of R. With the help of R, these companies are able to crunch data and process information, providing an essential backdrop for further analysis and data processing.

For more advanced processing like drug discovery, R is most widely used for performing pre-clinical trials and analyzing the drug-safety data. It also provides a suite for performing exploratory data analysis and vivid visualization tools to its users.

R is also popular for its *Bioconductor package* that provides various functionalities for analyzing the genomic data. R is also used for statistical modeling in the field of epidemiology, where data scientists analyze and predict the spread of diseases.

4. Social Media

For many beginners in Data Science and R, social media is a data playground. Sentiment Analysis and other forms of social media data mining are some of the important statistical tools that are used with R.

Social Media is also a challenging field for Data Science because the data prevalent on social media

websites is mostly unstructured in nature. R is used for social media analytics, for segmenting potential customers and targeting them for selling your products.

Furthermore, mining user sentiment is another popular category in social media analytics. With the help of R, companies are able to model statistical tools that analyze user sentiments, allowing them to improve their experiences.

SocialMediaMineR *is a popular R tool that can take multiple URLs and churn the popularity of their reach on social media*. Furthermore, companies use R to analyze the social media market and generate leads for the user.

5. E-Commerce

The e-commerce industry is one of the most important sectors that utilize Data Science. *R is one of the standard tools that is being used in e-commerce*.

Since these internet-based companies have to deal with various forms of data, structured and unstructured, as well as varying data sources like spreadsheets and databases (SQL & NoSQL), R proves to be an effective choice for these industries.

E-commerce companies use R for analyzing cross-selling products to their customers. In cross-selling, a customer is suggested an additional product that complements their original purchase. These types of suggestions and recommendations are best analyzed with the help of R.

Various statistical procedures like linear modeling are necessary to analyze the purchases made by the customers as well as in predicting product sales. Furthermore, companies use R for carrying out A/B testing analysis across the pages of their products.

Some more Applications of R

These are some more R applications, which you can use to make better decision making:

1. R is primarily used for ***descriptive statistics***. Descriptive statistics summarize the main features of the data. R is used for a variety of purposes in summary statistics like central tendency, measurement of variability, finding kurtosis and skewness.
2. R is most widely used for *exploratory data analysis*. R's most popular package **ggplot2** is considered to be one of the best visualization libraries due to its aesthetics and interactivity.
3. R is also used for analyzing *both discrete and continuous probability distributions*. For example, using the *ppois()* function, you can draw Poisson distribution. Similarly, with the help of *dbinom()* function, you can plot the binomial distribution.
4. R also allows hypothesis testing to validate statistical models.
5. You can find a correlation between the variables in R using the *lm()* function that is used for establishing linear regression as well as multivariable linear regression.
6. Using R, you can make use of the *tidyverse package* that is used for organizing data and data pre-processing.
7. R also provides an *interactive web application package called* **RShiny**. With this package, you can develop interactive visualizations that can be embedded on your web-pages.

8. Moreover, with the help of R, you can *develop predictive models* that make use of machine learning algorithms to find the occurrences of future events.