Assignment 1

Tanmay Pardeshi
43260 - R 10

Problem Statement : Download the open source software of your interest. Document the distinct features and functionality of the software platform. You may choose WEKA and R and Python

Python :

Python is a programming language that distinguishes itself from other programming languages by its flexibility, simplicity, and reliable tools required to create modern software. Python is consistent and is anchored on simplicity, which makes it most appropriate for machine learning. The Python programming language best fits machine learning due to its independent platform and its popularity in the programming community.

Features of Python:

- Python has a smooth learning curve. It is easy to learn.
- It is a readable language.
- Python is an interpreted language. It comes with the IDLE (Interactive Development Environment). This is an interpreter and follows the REPL structure (Read-Evaluate-Print-Loop). It executes and displays the output of one line at a time.
- It is a dynamically typed language. You don’t need to declare a data type while defining a variable. The interpreter determines this at runtime based on the types of the parts of the expression.
- Python is object-oriented but supports both functional and object-oriented programming. Everything in Python is an object.
- Open-source and it has one of the largest community support.
- Python is platform-independent. If you write a program, it will run on different platforms like Windows, Mac and Linux.

- Large standard library. The standard library is large and has many packages and modules with common and important functionality. You can also install packages from the PyPI (Python Package Index) if you want even more functionality.

R:

R is the most popular programming language for statistical modeling and analysis.It is heavily used in analyzing data that is both structured and unstructured. This has made R, the standard language for performing statistical operations. R allows various features that set it apart from other Data Science languages. It is a continuously evolving language which means that many cons will slowly fade away with future updates to R.

Features of R:

- R is an open-source software environment. It is free of cost and can be adjusted and adapted according to the user's and the project's requirements.
- R can produce static graphics with production quality visualizations and has extended libraries providing interactive graphic capabilities. This makes data visualization and data representation very easy.
- R has an open-source library which is supported by its growing number of users.
- R has a very comprehensive development environment meaning it helps in statistical computing as well as software development.
- R is an object-oriented programming language.
- R can be used to perform simple and complex mathematical and statistical calculations on data objects of a wide variety. It can also perform such operations on large data sets.
- R's markdown package is the only report generation package you will ever need when working with R. The markdown package can help produce web pages.
- R can be easily paired with other data processing and distributed computing technologies like Hadoop and Spark. It is possible to remotely use a Spark cluster to process large datasets using R.
- Vectors are the most basic data structure in R, and most other data structures are derived from vectors.

WEKA:

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization.

Features of WEKA:

- Platform independent
- Open source and free
- Different machine learning algorithms for data mining
- Easy to use
- Data preprocessing tools
- Flexibility for scripting experiments
- Graphical user interface