

Assignment 5

Study of Open Source Analytical Software

AIM

Study of platform for Implementation of Assignments. Download and install Python, and the open source software - WEKA and R. Document the distinct features and functionality of the software platform.

OBJECTIVE

To study -

- Concept of open source analytical software(WEKA and R).
- Concept of statistical analysis.
- Distinct features and functionality of open source software.

THEORY

INTRODUCTION TO PYTHON

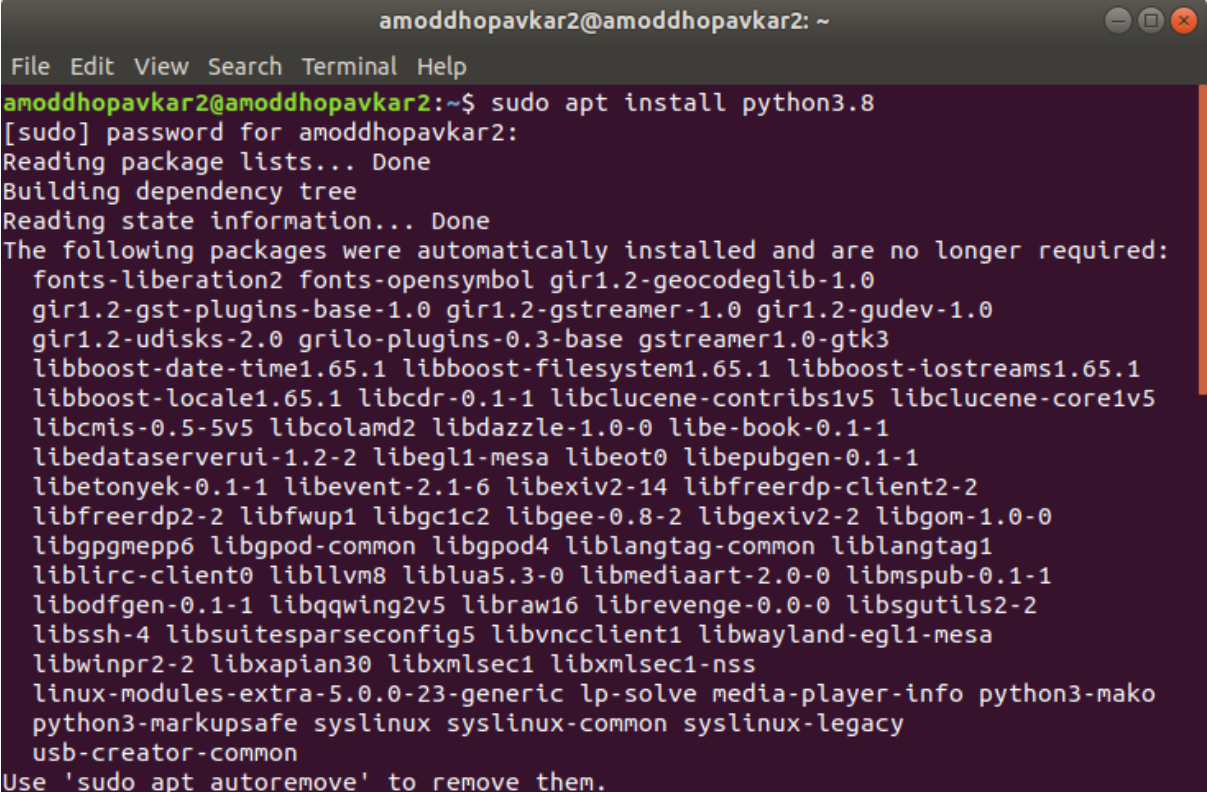
Python is an interpreted, high-level and general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace.

Features -

- Easy to Code. Python is a very developer-friendly language which means that anyone and everyone can learn to code it in a couple of hours or days.
- Open source and free.
- GUI support.
- Object Oriented approach.
- High-level language.
- Highly portable.
- Highly dynamic.
- Integrated by nature.

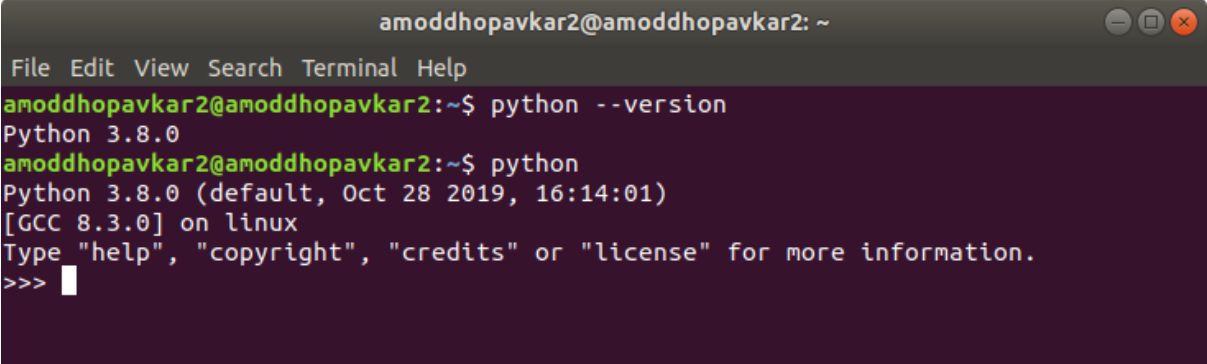
Installation -

- Install python with the command “sudo apt install python3.8”.



```
amoddhopavkar2@amoddhopavkar2: ~  
File Edit View Search Terminal Help  
amoddhopavkar2@amoddhopavkar2:~$ sudo apt install python3.8  
[sudo] password for amoddhopavkar2:  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following packages were automatically installed and are no longer required:  
  fonts-liberation2 fonts-opensymbol gir1.2-geocodeglib-1.0  
  gir1.2-gst-plugins-base-1.0 gir1.2-gstreamer-1.0 gir1.2-gudev-1.0  
  gir1.2-udisks-2.0 grilo-plugins-0.3-base gstreamer1.0-gtk3  
  libboost-date-time1.65.1 libboost-filesystem1.65.1 libboost-iostreams1.65.1  
  libboost-locale1.65.1 libcdr-0.1-1 libclucene-contribs1v5 libclucene-core1v5  
  libcmis-0.5-5v5 libcolamd2 libdazzle-1.0-0 libe-book-0.1-1  
  libedataserverui-1.2-2 libegl1-mesa libeot0 libepubgen-0.1-1  
  libetonyek-0.1-1 libevent-2.1-6 libexiv2-14 libfreerdp-client2-2  
  libfreerdp2-2 libfwup1 libgc1c2 libgee-0.8-2 libgexiv2-2 libgom-1.0-0  
  libgpgmepp6 libgpod-common libgpod4 liblangtag-common liblangtag1  
  liblirc-client0 libllvm8 liblua5.3-0 libmediaart-2.0-0 libmspub-0.1-1  
  libodfgen-0.1-1 libqqwing2v5 libraw16 librevenge-0.0-0 libsgutils2-2  
  libssh-4 libsuitesparseconfig5 libvncclient1 libwayland-egl1-mesa  
  libwinpr2-2 libxapian30 libxmlsec1 libxmlsec1-nss  
  linux-modules-extra-5.0.0-23-generic lp-solve media-player-info python3-mako  
  python3-markupsafe syslinux syslinux-common syslinux-legacy  
  usb-creator-common  
Use 'sudo apt autoremove' to remove them.
```

- Check the installed version and run the python shell using commands “python --version” and “python”.



```
amoddhopavkar2@amoddhopavkar2: ~  
File Edit View Search Terminal Help  
amoddhopavkar2@amoddhopavkar2:~$ python --version  
Python 3.8.0  
amoddhopavkar2@amoddhopavkar2:~$ python  
Python 3.8.0 (default, Oct 28 2019, 16:14:01)  
[GCC 8.3.0] on linux  
Type "help", "copyright", "credits" or "license" for more information.  
>>> █
```

INTRODUCTION TO R

R is a programming language and free software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis.

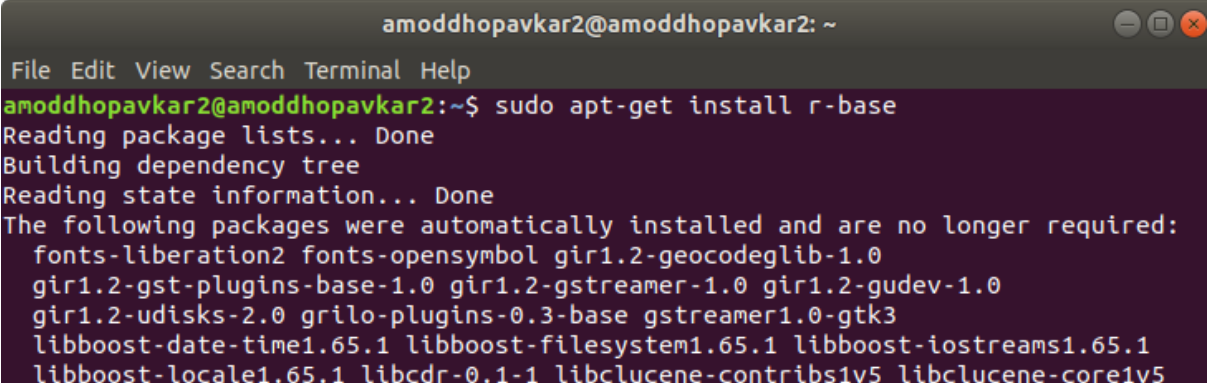
It was created by Ross Ihaka and Robert Gentleman (Hence the name 'R') at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team, of which Chambers is a member. R is named partly after the first names of the first two R authors and partly as a play on the name of S.R is a GNU project.

Features -

- R is a well-developed, simple and effective programming language which includes conditionals, loops, user defined recursive functions and input and output facilities.
- R has an effective data handling and storage facility.
- R provides a suite of operators for calculations on arrays, lists, vectors and matrices.
- R provides a large, coherent and integrated collection of tools for data analysis.
- R provides graphical facilities for data analysis and display either directly at the computer or printing at the papers.

Installation -

- Install R with the command "sudo apt-get install r-base".

A terminal window with a dark background and light text. The window title is 'amoddhopavkar2@amoddhopavkar2: ~'. The menu bar shows 'File Edit View Search Terminal Help'. The terminal output shows the command 'sudo apt-get install r-base' being executed. It reports 'Reading package lists... Done', 'Building dependency tree', and 'Reading state information... Done'. A list of packages that were automatically installed and are no longer required is displayed, including fonts-liberation2, fonts-opensymbol, gir1.2-geocodeglib-1.0, gir1.2-gst-plugins-base-1.0, gir1.2-gstreamer-1.0, gir1.2-gudev-1.0, gir1.2-udisks-2.0, grilo-plugins-0.3-base, gstreamer1.0-gtk3, libboost-date-time1.65.1, libboost-filesystem1.65.1, libboost-iostreams1.65.1, libboost-locale1.65.1, libcdr-0.1-1, libclucene-contribs1v5, and libclucene-core1v5.

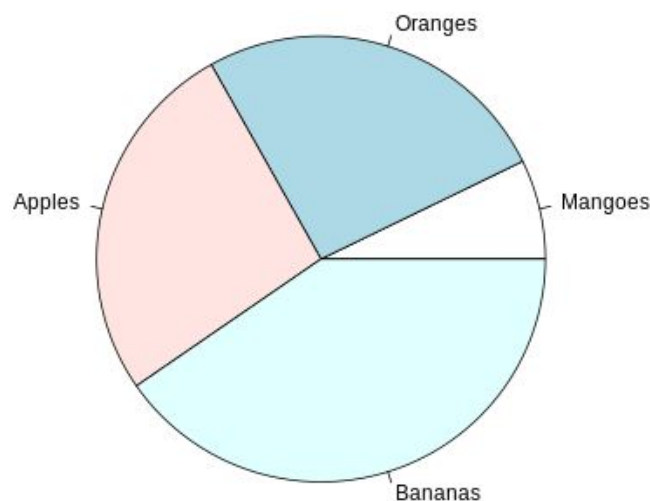
```
amoddhopavkar2@amoddhopavkar2: ~  
File Edit View Search Terminal Help  
amoddhopavkar2@amoddhopavkar2:~$ sudo apt-get install r-base  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following packages were automatically installed and are no longer required:  
  fonts-liberation2 fonts-opensymbol gir1.2-geocodeglib-1.0  
  gir1.2-gst-plugins-base-1.0 gir1.2-gstreamer-1.0 gir1.2-gudev-1.0  
  gir1.2-udisks-2.0 grilo-plugins-0.3-base gstreamer1.0-gtk3  
  libboost-date-time1.65.1 libboost-filesystem1.65.1 libboost-iostreams1.65.1  
  libboost-locale1.65.1 libcdr-0.1-1 libclucene-contribs1v5 libclucene-core1v5
```

- Type "R" on terminal/Command line to get command line for R programming.

```
amoddhopavkar2@amoddhopavkar2: ~  
File Edit View Search Terminal Help  
amoddhopavkar2@amoddhopavkar2:~$ R  
  
R version 3.4.4 (2018-03-15) -- "Someone to Lean On"  
Copyright (C) 2018 The R Foundation for Statistical Computing  
Platform: x86_64-pc-linux-gnu (64-bit)  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
Natural language support but running in an English locale  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

- Demo - Create a simple pie chart in R for the sale of various fruits.

```
> # demo by amoddhopavkar2  
> # A basic pie chart  
> x <- c(12, 43, 44, 67)  
> labels <- c("Mangoes", "Oranges", "Apples", "Bananas")  
>  
> # Name the file  
> png(file = "fruits.png")  
>  
> # Plot the chart  
> pie(x, labels)  
>  
> # Save the file  
> dev.off()  
null device
```



INTRODUCTION TO WEKA

WEKA is open source software under the GNU General Public License. The system is written using object oriented language Java. There are several different levels at which WEKA can be used. Weka provides implementations of state-of-the-art data mining and machine learning algorithms. WEKA contains modules for data preprocessing, classification, clustering and association rule extraction.

Features -

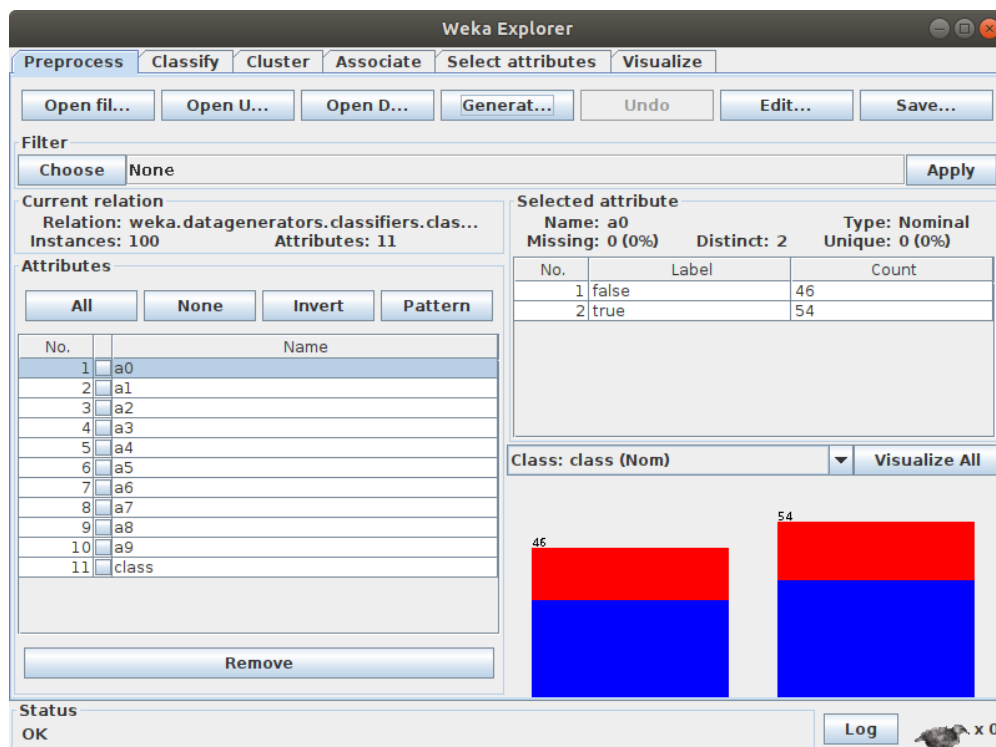
- Preprocess
- Classify
- Cluster
- Associate
- Select attributes
- Visualize

Installation -

- Install using the command “sudo apt install weka”

```
amoddhopavkar2@amoddhopavkar2: ~  
File Edit View Search Terminal Help  
amoddhopavkar2@amoddhopavkar2:~$ sudo apt install weka  
[sudo] password for amoddhopavkar2:  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done
```

- Demo



CONCLUSION

Downloaded Python and the open source softwares R-base, RStudio and WEKA, and studied the distinct features and functionality of these software platforms.