

and its associated notations for data analytics and visualization.

UNIT I:

Defining Data Science and Big data, Benefits and Uses, facets of Data, Data Science Process. History and Overview of R, Getting Started with R, R Nuts and Bolts

UNIT II:

The Data Science Process: Overview of the Data Science Process-Setting the research goal, Retrieving Data, Data Preparation, Exploration, Modeling, data Presentation and Automation. Getting Data in and out of R, Using reader package, Interfaces to the outside world.

UNIT III:

Machine Learning: Understanding why data scientists use machine learning-What is machine learning and why we should care about, Applications of machine learning in data science, Where it is used in data science, The modeling process, Types of Machine Learning-Supervised and Unsupervised.

UNIT IV:

Handling large Data on a Single Computer: The problems we face when handling large data, General Techniques for handling large volumes of data, Generating programming tips for dealing with large datasets. Case study- Predicting malicious URLs(This can be implemented in R).

UNIT V:

Sub setting R objects, Vectorised Operations, Managing Data Frames with the dplyr, Control structures, functions, Scoping rules of R, Coding Standards in R, Loop Functions, Debugging, Simulation