

MACHINE LEARNING BY USING R-PROGRAMMING



CURRICULUM

FUNDAMENTAL OF STATISTICS.

- Population and sample
- Descriptive and Inferential Statistics
- Statistical data analysis
- Variables
- Sample and Population Distributions
- Interquartile range
- Central Tendency
- Normal Distribution
- Skewness.
- Boxplot
- Five Number Summary
- Standard deviation
- Standard Error
- Empirical Formula
- central limit theorem
- Estimation
- Confidence interval
- Hypothesis testing
- p-value
- Scatterplot and correlation coefficient
- Standard Error
- Scales of Measurements and Data Types
- Data Summarization
- Visual Summarization
- Numerical Summarization
- Outliers & Summary

Module 1- Introduction to Data Analytics

Objectives:

- This module introduces you to some of the important keywords in R like Business Intelligence, Business
- Analytics, Data and Information. You can also learn how R can play an important role in solving complex analytical problems.
- This module tells you what is R and how it is used by the giants like Google, Facebook, etc.
- Also, you will learn use of 'R' in the industry, this module also helps you compare R with other software
- in analytics, install R and its packages.

Topics:

Business Analytics, Data, Information

- Understanding Business Analytics and R
- Compare R with other software in analytics
- Install R
- Perform basic operations in R using command line

Module 2- Introduction to R programming

Starting and quitting R

Recording your work

- Basic features of R.
- Calculating with R
- Named storage
- Functions
- R is case-sensitive
- Listing the objects in the workspace
- Vectors
- Extracting elements from vectors
- Vector arithmetic
- Simple patterned vectors
- Missing values and other special values
- Character vectors Factors
- More on extracting elements from vectors
- Matrices and arrays
- Data frames
- Dates and times

NOTE:-

- Assignments with Datasets

Import and Export data in R

- Importing data in to R
- CSV File
- Excel File
- Import data from text table

Topics

- Variables in R
- Scalars
- Vectors
- R Matrices
- List
- R – Data Frames
- Using c, Cbind, Rbind, attach and detach functions in R
- R – Factors
- R – CSV Files
- R – Excel File

NOTE:-

- Assignments
- Business Scenerio/Group Discussion.

R Nuts and Bolts:-

- Entering Input. – Evaluation- R Objects- Numbers- Attributes- Creating Vectors- Mixing Objects- Explicit Coercion- Summary- Names- Data Frames.

Module 3- Managing Data Frames with the dplyr package

- The dplyr Package
- Installing the dplyr package
- select()
- filter()
- arrange()
- rename()
- mutate()
- group_by()
- %>%

NOTE-:

- Assignments
- Business Scenerio/Group Discussion.

Module 4- Loop Functions

- Looping on the Command Line
- lapply()
- sapply()
- tapply()
- apply()

NOTE-:

- Assignments
- Business Scenerio/Group Discussion.

Module 5- Data Manipulation in R Objectives:

- In this module, we start with a sample of a dirty data set and perform Data Cleaning on it, resulting in a data set, which is ready for any analysis.
- Thus using and exploring the popular functions required to clean data in R.

Topics

- Data sorting
- Find and remove duplicates record
- Cleaning data
- Merging data

Statistical Plotting-:

- Bar charts and dot charts
- Pie charts
- Histograms
- Box plots
- Scatterplots
- QQ plots

NOTE:-

- Assignments with Datasets

Objectives:

- Control Structure Programming with R
- The for() loop
- The if() statement
- The while() loop
- The repeat loop, and the break and next statements
- Apply
- Sapply
- Lapply

NOTE:-

- Assignments with Datasets

Factors

- Using Factors
- Manipulating Factors
- Numeric Factors
- Creating Factors from Continuous Variables
- Convert the variables in factors or in others.

Reshaping

- Data Modifying
- Data Frame Variables
- Recoding Variables
- The recode Function

- Reshaping Data Frames
- The reshape Package

NOTE:-

- Assignments with Datasets

Explore many algorithms and models:

- Popular algorithms: Classification, Regression, Clustering, and Dimensional Reduction.
- Popular models: Train/Test Split, Root Mean Squared Error, and Random Forests. Get ready to do more learning than your machine!

Module 6 - Machine Learning vs Statistical Modeling & Supervised vs Unsupervised Learning

- Machine Learning Languages, Types, and Examples
- Machine Learning vs Statistical Modelling
- Supervised vs Unsupervised Learning
- Supervised Learning Classification
- Unsupervised Learning

Module 7 - Supervised Learning I

- K-Nearest Neighbors
- Decision Trees
- Random Forests
- Reliability of Random Forests
- Advantages & Disadvantages of Decision Trees

Module 8 - Supervised Learning II

- Regression Algorithms
- Model Evaluation
- Model Evaluation: Overfitting & Underfitting
- Understanding Different Evaluation Models

Module 9 - Unsupervised Learning

- K-Means Clustering plus Advantages & Disadvantages
- Hierarchical Clustering plus Advantages & Disadvantages
- Measuring the Distances Between Clusters - Single Linkage Clustering
- Measuring the Distances Between Clusters - Algorithms for Hierarchy Clustering
- Density-Based Clustering

Module 10 - Dimensionality Reduction & Collaborative Filtering

- Dimensionality Reduction: Feature Extraction & Selection
- Collaborative Filtering & Its Challenges

Module 11 - Classification:-

- An Overview of Classification.
- Why Not Linear Regression?
- Logistic Regression
- The Logistic Model
- Estimating the Regression Coefficients
- Making Predictions
- Logistic Regression for >2 Response Classes
- Lab: Logistic Regression.
- The Stock Market Data
- Logistic Regression

NOTE:-

- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Module 12 - Tree-Based Methods:-

- The Basics of Decision Trees
- Regression Trees
- Classification Trees
- Trees Versus Linear Models
- Advantages and Disadvantages of Trees
- Bagging, Random Forests, Boosting
- Bagging

- Random Forests
- Lab: Decision Trees
- Fitting Classification Trees
- Fitting Regression Trees

NOTE-:

- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Module 13 - Time Series & Forecasting-:

- Time series
- Estimating and Eliminating the Deterministic Components if they are present in the Model.
- Estimating and Eliminating Seasonality if it is present in the Model
- Modeling the Remainder using Auto Regressive Moving Average (ARMA) Models
- Identify 'order' of the ARMA model
- 'Forecast' or Predict for Future Values
- Practise on R

NOTE-:

- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Module 14 - Support Vector Machines – Outline

- Understand when the Support Vector family of methods are an appropriate method of analysis.
- Understand what a hyperplane is and how they are used with the Support Vector methods.
- Identify the differences between Maximal Margin Classifiers, Support Vector Classifiers, and Support Vector Machines.
- Know how each of the algorithms determines the best separating hyperplane.
- Distinguish between hard and soft margins and when each is to be used.
- Know how to extend the method for nonlinear cases.

NOTE-:

- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Module 15 - Principal Component Analysis – Outline

- Understand what principal components are and when principal component analysis is appropriate.
- Describe eigenvalues and eigenvectors and how they are used to calculate principal components.
- Understand loading and loading vectors.
- Know how to decide how many principal components to use in the analysis.
- Be able to use principal component analysis for regression.

NOTE-:

- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Partners : 				
---	---	---	---	---



E-mail: info@ducatindia.com
Visit us: www.ducatinidia.com
www.facebook.com/ducateducation

NOIDA

A-43 & A-52, Sector-16,
 Noida - 201301, (U.P.) INDIA
 Ph. : 0120-4646464
 Mb. : 09871055180

GURGAON

1808/2, 2nd floor old DLF,
 Near Honda Showroom,
 Sec.-14, Gurgaon (Haryana)
 Ph. : 0124-4219095-96-97-98
 Mb. : 09873477222-333

GREATER NOIDA

F 205 Neelkanth Plaza Alpha 1
 commercial Belt Opposite to Alpha
 Metro Station Greater Noida
 Ph. : 0120-4345190-91-92 to 97
 Mb. : 09899909738, 09899913475

GHAZIABAD

1, Anand Industrial Estate,
 Near ITS College, Mohan Nagar,
 Ghaziabad (U.P)
 Ph. : 0120-4835400...98-99
 Mb. : 09810831363 / 9818106660
 : 08802288258 - 59-60

FARIDABAD

SCO-32, 1st Floor, Sec.-16,
 Faridabad (HARYANA)
 Ph. : 0129-4150605-09
 Mb. : 09811612707