Add on course - Data Science & Analytics (with Introduction to Tableau)

Miranda House, University of Delhi

Audience

This program is designed for enthusiasts considering that they have some basic knowledge of statistics are looking to acquire working knowledge of Machine Learning, Deep Learning concepts using R and R Studio and important concepts in Tableau.

Resource Person(s)

The faculties are arranged from Corporates and Academia with excellent hands-on experience of working on R and its use in the industry in multiple domains such as Healthcare, FMCG, Telecom, Education, F&B and more.

Training Content

The course is composed of 30 hours of Theory + Practical learning towards R Programming & Tableau, 10 hours of dedicated hands-on practice sessions and 4 hours towards the capstone project presentation and discussion. Total 44 hours. Each session is of 2.5 hours duration.

MODULE 1: BASIC STATISTICS BRUSHUP

- Central Tendency – Mean, Mode, Media, Variance
- Distributions
- Hypothesis Testing
- Confidence Interval
- Correlation
- Linear Regression

MODULE 2: OVERVIEW OF R

- Defining the R Project
- Getting Familiar with R Environment

MODULE 3: PROGRAMMING IN R PART 1

- R Nuts and Bolts – Essentials, Entering Input, Evaluation, R Objects, Numbers, Attributes, Creating Vectors, MixingObjects, Explicit Coercion, Matrices, Lists, Factors, Missing Values, Data Frames, Names, Summary

- Getting Data In and Out of R - Reading Data Files with read.table(), Reading Larger Datasets with read.table(),UsingTextual and Binary formats for Storing Data, Interfaces to Outside World, Reading Lines of a Text File, Reading Data from Internet and URL Connections

MODULE 4: PROGRAMMING IN R PART 2

- Subsetting R Objects - Subsetting a Vector, Matrix, Lists
- Vectorized Operations
- Dates and Times – Dates and Times in R, Operations on Dates and Times
- Control Structures - if-else, for Loops, Nested for Loops, while Loops, repeat Loops, next, break
- Apply Family of Functions – lapply, sapply, apply, tapply, split, mapply
- Sampling in R – Simulation, Random Sampling

MODULE 5: EXPLORATORY DATA ANALYSIS

- Basic distribution of data
- Summarization: Measures of Central Tendency, Dispersion, Skewness and Kurtosis
- Data Visualization: Histogram/Bar Chart, Box Plot, Stem and Leaf Display, Pairwise Scatter Plots
- Missing Value, Outlier Detection
- Testing of Normality: Histogram, QQ Plot, KS Test and SW Test
- Correlation Analysis

MODULE 6: STATISTICAL INFERENCE

- Parameter Estimation
- Non – Parametric Estimation
- Parametric Testing of Hypothesis I – Testing of Hypothetical Value of Population Mean and Variance
- Parametric Testing of Hypothesis II – Testing for Equality of two Population Means and Variances, SeveralPopulation Mean
- Non – Parametric Testing of Hypothesis I – Testing for Hypothetical value of population median, Testing forEquality of Two and Several Populations
- Non – Parametric Testing of Hypothesis II – Testing for Goodness of fit, Testing for Independence of Attributes

MODULE 7: LINEAR REGRESSION ANALYSIS

- Model Building - Fitting a Linear Regression Model, Testing the significance of individual regressors and overallregression, Goodness of the Model: R Square and Adjusted R Square.
- Multicollinearity – Problems and its Consequences, Detection and Removal of Multicollinearity using CorrelationAnalysis, Variance Inflation Factors (VIFs)
- Parsimonious Modelling or Model Selection - Forward Selection, Backward Elimination, Stepwise Selection
- Validation of Assumptions and Residual Analysis - Linearity of Regression, Autocorrelation, Heteroscedasticity,Normality of Errors, Outliers Detection

MODULE 8: LOGISTIC REGRESSION

- Fitting a Logistic Regression Model
- Testing the Significance of Individual Regressors and Overall Regression
- Goodness of the Model: Confusion Matrix
- Sensitivity and Specificity
- Odds Ratio
- Multiclass Classification

MODULE 9: RANDOM FOREST

MODULE 10: UNSUPERVISED LEARNING – CLUSTER AND FACTOR ANALYSIS

MODULE 11: TABLEAU

- Basic and Advanced Data Preparation
- Data Visualization: Bar Charts, Line Charts, Maps, Scatterplots
- Advanced Story-telling dashboard design

MODULE 12: DEEP LEARNING

- Introduction to ANN (Artificial Neural Networks)
- Building ANN in R
- Introduction to CNN (Convolutional Neural Networks)
- Building CNN in R
- Textual data Introduction and Application in R using RNN (Recurrent Neural Networks)