

Data Science using R Programming

Module 1: Introduction to R programming

1. History of R-language
2. Why to learn R-language
3. Importance of R-language
4. Installation and setup Environment
5. Packages interfaces and library

Module 2: R programming essentials

1. Expressions and Operations
2. Data Types and Data Structures- Vectors, Factors,Matrix, Dataframes,Lists
3. Vector Basics
4. Vector Operations
5. Vector Indexing and Slicing
6. Matrix Operations

Module 3: Fundamentals of R Language

1. Data Frame Indexing and Selection
2. Operations on Data Frame
3. CSV Files with R
4. Operators
5. Conditional Statements
6. Loops & Functions
7. Built-in R Features & Apply
8. Dates and Timestamps

Module 4: Data Visualization and Data Manipulation

1. Understanding & Working with Graph Libraries.
2. Overview of ggplot2
3. Histograms
4. Scatterplots
5. Bar Plot
6. Boxplots
7. 2 Variable Plotting
8. Sorting, Concatenation of Datasets

Module 5: Testing of Hypothesis

1. Concept of Hypothesis.
2. Null Hypothesis
3. Alternative Hypothesis
4. Type-I error
5. Type-II error
6. Level of Significance
7. Confidence Intervals
8. Parametric Tests and Non Parametric Tests
9. One Sample T test
10. Two Independent Sample T test
11. Paired Sample T test
12. Chi-square Test for Independence of Attributes

Module 6: Exploratory Factor Analysis

1. Principal Component Analysis
2. Concept of Communalities
3. Eigen Values and Eigen Vectors
4. Correlation Matrix check and KMO-MSA check
5. Factor loading Matrix
6. Diagrammatic Representation of Factors
7. Problems of Factor Loadings and Solutions

-----Supervised Machine Learning Algorithm-----

Module 7: Machine Learning with R-language

1. Introduction to Machine Learning
2. Data Munging in R
3. Cyclical vs Seasonal Analysis

Module 8: Analysis of Variance

1. One Way Anova
2. Two Way Anova

Module 9: Linear Regression Using R

1. Concept of Linear Regression.
2. Important features of a Straight line.
3. Method of least Square.
4. Assumptions of Classical Linear Regression Model
5. Understanding the Goodness of Fit
6. Test of Significance of the Estimated parameters.
7. Concept of multicollinearity
8. Concept of VIF
9. Concept of AutoCorrelation
10. Practical Application of Linear Regression using R.

Module 10: Logistic Regression Using R

1. Concept of Logistic Regression .
2. Differences between Linear Regression and Logistic Regression.
3. Logistic Regression Model.
4. ODDS AND ODDS RATIO-Mathematical Concepts
5. Concept of Concordant Pairs, Discordant Pairs, Tied Pairs.
6. Confusion Matrix and its Measures
7. Determining the Cut-Point Probability Level.
8. Receiver Operating Characteristic Curves
9. Practical Application of Logistic Regression using R.

Module 11: Time Series Analysis

1. Concept of Time Series and its Applications
2. Assumptions of Time Series Analysis
3. Components of Time Series
4. Smoothing techniques
5. Stationarity
6. Random Walk
7. ARIMA Forecasting
8. Box Jenkins Technology
9. Merits and Demerits of BJ Technology

-----Unsupervised Machine Learning Algorithm-----

Module 12: Decision tree and Clustering

1. Concept of Decision Tree
2. Decision Tree Application using R.
3. Concept of K-Means Clustering.
4. Types of Cluster Analysis.
5. Concept of Linkage.
6. Ward's Minimum Variance Criteria.
7. Clustering related Statistics-Semi-Partial R-Square, R Square
8. Graphical Representation of Cluster Analysis
9. Practical Application of Clustering using R.

-----Prescriptive Analytics-----

Module 13: Text Mining and Sentiment Analysis

1. Concept of Text Mining and Sentiment Analysis
2. Concept of Stopwords
3. Practical Application of Text Mining and Sentiment Analysis

Module 14: Market Basket Analysis

1. Concept of Market Basket Analysis
2. Measures of Market Basket Analysis-Support, lift, Confidence
3. Advantages of Market Basket Analysis
4. Practical Application of Market Basket Analysis.