



SYLLABUS OF ARTIFICIAL INTELLIGENCE WITH MACHINE LEARNING

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
- Emperical 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



B-8, SECTOR-2
NEXT TO SECTOR 15
METRO STATION, NOIDA
UP-201301



+91-9891740056
+91-9968389880
+91-120-4206078
Whatsapp: 9310785056



TRAINING SYLLABUS

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



**B-8, SECTOR-2
NEXT TO SECTOR 15
METRO STATION, NOIDA
UP-201301**



**+91-9891740056
+91-9968389880
+91-120-4206078
Whatsapp: 9310785056**



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

3. 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

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





MODULE 6- Data Management-I:-

- Understand Data Using summary() Function
- Check Missing Observations
- Extracting Complete Cases
- Handling Missing Values (Blank Entries) While Importing
- Subset Using Selected Rows
- Subset Using Selected Columns
- Selected Rows for Selected Columns

MODULE 7- Subset Function:-

- **Condition on observations**
- Condition on variable names
- Condition on observations and variable names
- Sub-setting with 'Not Equal To' Operator

Module 8- Data Management-II:-

- Data Sorting in R (Ascending)
- Data Sorting in R (Descending)
- Data Sorting in R (Using Factor Variable)
- Sorting Data by Multiple Variables

Module 9- Data Management-III (Modifying Data):-

- **Change Variable Names & Content Using fix()**
- Change Variable Names Using Index
- Change Variable Names Using setnames()
- Derive a New Variable
- Remove Rows from a Dataframe
- Remove Columns from a Dataframe





Module 10- Descriptive Statistics:-

- Dissect Data with Quantiles
- Quantiles in R
- Box-Whisker Plot
- Box-Whisker Plot in R
- Display Observation Number on Outliers

Module 11- Descriptive Statistics III Bivariate Relationship:-

- Scatter Plot
- Pearson's correlation coefficient
- Simple Linear Regression
- Using R for correlation and simple regression

Module 12- Statistical Learning:-

- What Is Statistical Learning?
- Why Estimate f ?
- How Do We Estimate f ?
- The Trade-Off Between Prediction Accuracy and Model Interpretability
- Supervised Versus Unsupervised Learning
- Regression Versus Classification Problems
- Assessing Model Accuracy





Module 13- Basics of Statistics & Linear & Multiple Regression

- This module touches the base of Descriptive and Inferential Statistics and Probabilities & 'Regression Techniques'.
- Linear and logistic regression is explained from the basics with the examples and it is implemented in R using two case studies dedicated to each type of Regression discussed.
- Assessing the Accuracy of the Coefficient Estimates.
- Assessing the Accuracy of the Model.
- Estimating the Regression Coefficients.
- Some Important Questions
- Lab: Linear Regression.
 - i. Libraries .
 - ii. Simple Linear Regression
 - iii. Multiple Linear Regression
 - iv. Interaction Terms
 - v. Qualitative Predictors
 - vi. Writing Functions

NOTE:-

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

Module 14- 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



**B-8, SECTOR-2
NEXT TO SECTOR 15
METRO STATION, NOIDA
UP-201301**



**+91-9891740056
+91-9968389880
+91-120-4206078
Whatsapp: 9310785056**



NOTE:-

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

Module 15- Variance Inflation Factor:-

- Introduction
- Multicollinearity.
- **How we can detect the multicollinearity.**
- Effects of multicollinearity
- Lab: VIF
 - i. **Mutiple Datasets.**
 - ii. Applications.
 - iii. Reduce the features.

NOTE:-

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

Correlation

- Types of Correlation
- Properties of Correlation
- Methods of Calculating Correlation

Module 16- Best Model Selection:-

- Subset Selection
- Best Subset Selection
- Stepwise Selection
- Choosing the Optimal Model
- Lab 1: Subset Selection Methods
- Best Subset Selection
- Forward and Backward Stepwise Selection
- Choosing Among Models Using the Validation Set Approach and Cross-Validation

NOTE:-



**B-8, SECTOR-2
NEXT TO SECTOR 15
METRO STATION, NOIDA
UP-201301**



**+91-9891740056
+91-9968389880
+91-120-4206078
Whatsapp: 9310785056**



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

Module-17-Machine Learning with Artificial Intelligence

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

Module1 8 - UnSupervised Learning with Artificial Intelligence

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

Module1 9 - Supervised Learning II

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

Module 20 - 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 21- Artifical Neural Networks:-

- What is an Artificial Neural Network?
- What are Artificial Neural Networks used for?
- What are neurons





MACHINE LEARNING BY USING R-PROGRAMMING

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
- 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



**B-8, SECTOR-2
NEXT TO SECTOR 15
METRO STATION, NOIDA
UP-201301**



**+91-9891740056
+91-9968389880
+91-120-4206078
Whatsapp: 9310785056**



R-PROGRAMMING

TRAINING SYLLABUS

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



B-8, SECTOR-2
NEXT TO SECTOR 15
METRO STATION, NOIDA
UP-201301



+91-9891740056
+91-9968389880
+91-120-4206078
Whatsapp: 9310785056



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

3. Import and Export data in R

- Importing data in to R
- CSV File
- Excel File





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

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()
- %>%





Module 4- Loop Functions

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

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



B-8, SECTOR-2
NEXT TO SECTOR 15
METRO STATION, NOIDA
UP-201301



+91-9891740056
+91-9968389880
+91-120-4206078
Whatsapp: 9310785056



Statistical Plotting:-

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

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

MODULE 6- Data Management-I:-

- Understand Data Using summary() Function
- Check Missing Observations
- Extracting Complete Cases
- Handling Missing Values (Blank Entries) While Importing
- Subset Using Selected Rows
- Subset Using Selected Columns
- Selected Rows for Selected Columns





MODULE 7- Subset Function-:

- Condition on observations
- Condition on variable names
- Condition on observations and variable names
- Sub-setting with 'Not Equal To' Operator

Module 8- Data Management-II-:

- Data Sorting in R (Ascending)
- Data Sorting in R (Descending)
- Data Sorting in R (Using Factor Variable)
- Sorting Data by Multiple Variables

Module 8- Data Management-III (Modifying Data)-:

- Change Variable Names & Content Using fix()
- Change Variable Names Using Index
- Change Variable Names Using setnames()

• **Derive a New Variable**

- Remove Rows from a Dataframe
- Remove Columns from a Dataframe





Module 9- Descriptive Statistics:-

- Dissect Data with Quantiles
- Quantiles in R
- Box-Whisker Plot
- Box-Whisker Plot in R
- Display Observation Number on Outliers

Module 10- Descriptive Statistics III Bivariate Relationship:-

- Scatter Plot
- Pearson's correlation coefficient
- Simple Linear Regression
- Using R for correlation and simple regression

Module 11- Basics of Statistics & Linear & Multiple Regression

- This module touches the base of Descriptive and Inferential Statistics and Probabilities & 'Regression Techniques'.
- Linear and logistic regression is explained from the basics with the examples and it is implemented in R using two case studies dedicated to each type of Regression discussed.
- Assessing the Accuracy of the Coefficient Estimates.
- Assessing the Accuracy of the Model.
- Estimating the Regression Coefficients.
- Some Important Questions
- Lab: Linear Regression.
 - i. Libraries .
 - ii. Simple Linear Regression
 - iii. Multiple Linear Regression
 - iv. Interaction Terms
 - v. Qualitative Predictors
 - vi. Writing Functions

NOTE:- Assignments with Different Datasets.

- Business Scenerio/Group Discussion





Module 8- 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 9- Variance Inflation Factor:-

- Introduction
- Multicollinearity.
- **How we can detect the multicollinearity.**
- Effects of multicollinearity
- Lab: VIF
 - i. **Mutiple Datasets.**
 - ii. Applications.
 - iii. Reduce the features.

NOTE:-

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





Correlation

- Types of Correlation
- Properties of Correlation
- Methods of Calculating Correlation

Module 10- Best Model Selection:

- Subset Selection
 - Best Subset Selection
 - Stepwise Selection
 - Choosing the Optimal Model
 - Lab 1: Subset Selection Methods
 - Best Subset Selection
-
- Forward and Backward Stepwise Selection
 - Choosing Among Models Using the Validation Set Approach and Cross-Validation

NOTE:-

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

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!



B-8, SECTOR-2
NEXT TO SECTOR 15
METRO STATION, NOIDA
UP-201301



+91-9891740056
+91-9968389880
+91-120-4206078
Whatsapp: 9310785056



Module-11-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

Module12 - Supervised Learning I

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

Module13 - Supervised Learning II

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

Module 14 - 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 15 - Dimensionality Reduction & Collaborative Filtering

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

Module 16- 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 17- 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-18:-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-19-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.

