



Social Prachar

# Data Science

Using R, Python, Statistics, Machine Learning, Deep Learning (using Tensorflow and Keras), Natural Language Processing, Computer Vision

---

<b>I. How to Be a Successful Data Scientist / AI Specialist</b>	<b>2</b>
<b>II. Introduction to Data Science and How to Be a Good Data Scientist:</b>	<b>2</b>
<b>1. SQL (Basics)</b>	<b>2</b>
<b>2. R Programming</b>	<b>3</b>
<b>3. Statistical Analysis Using R and Python</b>	<b>4</b>
<b>4. Understanding Python</b>	<b>5</b>
<b>5. Deep Dive into Statistics</b>	<b>7</b>
<b>6. Concepts of Data</b>	<b>8</b>
<b>7. Data Exploration or Exploratory Data Analysis (EDA)</b>	<b>8</b>
<b>8. Data Preparation and Transformation</b>	<b>9</b>
<b>9. Machine Learning</b>	<b>9</b>
<b>10. Feature Engineering</b>	<b>11</b>
<b>11. Deep Learning</b>	<b>12</b>
<b>12. Natural Language Processing/Text mining</b>	<b>13</b>
<b>Tools</b>	<b>14</b>
<b>13. Tensor Flow</b>	<b>14</b>
<b>14. Hadoop</b>	<b>14</b>
<b>15. Keras</b>	<b>14</b>
<b>16. Spark</b>	<b>14</b>
<b>III. Case Studies</b>	<b>14</b>
<b>IV. Projects (3 -15)</b>	<b>14</b>
<b>V. Monthly Tests</b>	<b>14</b>
<b>VI. Mock interviews</b>	<b>14</b>
<b>VII. HR Activities</b>	<b>14</b>
<b>VIII. Certification</b>	<b>14</b>

## Highlights

- ★ Topic Wise Assignments Every Week
- ★ Quick Recap & Review sessions on each Wednesday
- ★ Monthly Mock Tests
- ★ Workshops
- ★ Final Mock Test at the end of the course
- ★ 5 Projects to Cover All The Topics

## I. How to Be a Successful Data Scientist / AI Specialist

1. Train Your Brain
2. Methodology to Understand the Concepts Faster and Not Forget
3. Prepare Your Own PDF Material from IPython Notebooks
4. Plagiarism
5. Saving Your Work
6. Error Debugging

## II. Introduction to Data Science and How to Be a Good Data Scientist:

1. What Is Data Science?
2. Why We Need Data Science?
3. What Data Scientist Do?
4. Roles in Data Science
5. Data Science Future
6. Skills Required for Data Science?
7. Math Requirement?
8. Statistics Requirement?
9. Knowledge Resources

### 1. SQL (Basics)

- SQL Basics
- Retrieve Data from SQL Database

## 2. R Programming

1. R Basics, background
2. Comprehensive R Archive Network
3. Demo of Installing R On windows from CRAN Website
4. Installing R Studios on Windows OS
5. Setting Up R Workspace.
6. Getting Help for R-How to use help system
7. Installing Packages – Loading And Unloading Packages
- 8. Getting familiar with basics**
9. Operators in R – Arithmetic,Relational,Logical and Assignment Operators
10. Variables,Types Of Variables,Using variables Conditional statements, ifelse(), switch
11. Loops: For Loops,While Loops,Using Break statement,Switch
- 12. The R Programming Language- Data Types**
13. creating data objects from the keyword.
14. How to make different type of data objects.
15. Types of data structures in R
16. Arrays And Lists- Create Access the elements
17. Vectors – Create Vectors,Vectorized Operations,Power of Vectorized Operations  
Matrices- Building the first matrices,Matrix Operations,Subsetting,visualising subset
18. Data Frames- create and filter data frames,Building And Merging data frames.
- 19. Functions And Importing data into R**
20. Function Overview – Naming Guidelines
21. Arguments Matching,Function with Multiple Arguments
22. Additional Arguments using Ellipsis,Lazy Evaluation Multiple Return Values Function as  
Objects,Anonymous Functions
23. Importing and exporting Data into R- importing from files like excel,csv and minitab.
24. Import from URL and excel Files
25. Import from database.
- 26. Data Descriptive**
27. Statistics,Tabulation,Distribution
28. Summary Statistics for Matrix Objects. apply() Command. Converting an Object into a  
Table
29. Histograms, Stem and Leaf Plot, Density
30. Normal Distribution

### **31. Graphics in R – Types of graphics**

32. Bar Chart,Pie Chart,Histograms- Create and edit.

33. Box Plots- Basics of Boxplots- Create and Edit Visualisation in R using ggplot2.

34. More About Graphs: Adding Legends to Graphs,Adding Text to Graphs, Orienting the Axis Label.

## **3. Statistical Analysis Using R and Python**

35. Data Types

36. Exploratory Data Analysis ( Mean , media , Mode , Range, SD , Variance, Skewness ,

37. Kurtosis

38. Data Visualizations

39. Graphical Representation of Various Charts (Bar Plot, Box Plot etc...)

40. Probability Distribution

41. Confidence Interval

42. Z Test

43. T Test

44. Anova

45. Hypothesis Testing ( Type I and TYPE II ERRORS)

### **Imputation Technique & Regression Technique**

1. Scatter Plot

2. Regression Analysis

3. Simple Linear Regression with R

4. Multiple Linear Regression with R

5. Multiple Logistic Regression with R

### **Data Exploration**

- Data Mining – Unsupervised Learning using R
- Data Mining – supervised Learning using R
- Dimension Reduction – Principal Component Analysis Using R
- Association Rules using R

## 4. Understanding Python

- Getting started with Python
- Python Overview
- About Interpreter languages
- Advance /Disadvantages of Python
- Starting Python
- Interpreter Path
- Using the interpreter
- Running a Python Script
- Keywords
- Built-In Function
- String Different Literals
- Math Operators and Expressions
- Writing on the screen
- String formatting
- Command line parameters and Flow control
- Numbers and Math's
  - Arithmetic, Floats and Modulo
- Ordering Operations using
- Variables and Inputs
  - Creating Variables
  - Input functions
- Dictionary
- Conditional Statements
  - If
  - Else
  - Elif
- Loops
  - While loop
  - For loop
- Reading and Writing
- Modules and Packages

### Sequence and File operations

- string
- Lists
- Tuples
- Set

- Dictionary
- Indexing and Slicing
- Iterating through a sequence
- Functions for all sequence
- Using Enumerate()
- Operators and Keywords for sequence
- The xrange() Function
- List comprehensions
- Generator expression

### **Deep Dive – Function sorting Error and Exception Handling**

- Functions
- Functions parameters
- Variable scope and Returning values. sorting
- Alternative Keys
- Lambda Functions
- Sorting collection of collections
- Sorting dictionaries
- Sorting list in place
- Errors and Exceptions handling
- Handling Multiple Exceptions
- The standard Exception hierarchy
- Using Modules
- The import statement Module search path
- Package installation ways

### **Regular Expressionist's Packages and Object oriented programming in python**

- The Sys Module
- Interpreter iteration
- STDIO
- Launching external programs
- Paths Directories and filenames
- Walking Directory tree
- Math function
- Random numbers
- Zipped Archives
- Introduction to Python class
- Defining classes



- Initializes
- Instance Methods
- Properties
- Class Methods and Data Static Methods
- Private Methods and Inheritance
- Module Aliases and Regular Expressing

## Libraries

- Crash course on **Pandas** –For Data Manipulation
- Crash course on **Numpy** – For Array-Processing
- Crash course on **Matplotlib** – For Visualization
  
- Scipy
- Scikit-Learn
- Keras
- Seaborn
- Cufflinks
- NLTK

## Debugging, Databases and project skeletons

- Debugging
- Dealing with Errors

## Linear Algebra

- Introduction to Linear Algebra
- Applications of Linear Algebra

## 5. Deep Dive into Statistics

### Understand Your Data Using Descriptive Statistics

- Measure of central tendency
  - a. Mean
  - b. Median
  - c. Mode
- Measures of Dispersion
  - a. Range quartile

- b. Inter-quartile rang
  - c. Variance and Standard deviation
- Measures to describe shape of distribution
  - a. **Skewness**
  - b. **Kurtosis**
- Probability Distributions
- Sampling Distributions and Confidence Intervals

#### Understand Your Data Using **Inferential Statistics**

- a. Hypothesis testing
- b. Chi Square Statistic and Contingency Tables
- c. T-test or ANOVA
- d. Correlation

## 6. Concepts of Data

- Collecting data from different sources
- Analyzing data
- Data preprocessing
- Data munging
- Data mining
- Data manipulation
- Data visualization
- Feature Selection
- Feature Scaling
- Dimensionality reduction

## 7. Data Exploration or Exploratory Data Analysis (EDA)

- a. Variable Identification
- b. Univariate Analysis (Exploring Individual Features)
- c. Bivariate Analysis (Exploring Two or Multi-Feature Relationships)
- d. Covariance & Correlation
- e. Multicollinearity
- f. Dimensionality Reduction using PCA

## 8. Data Preparation and Transformation

- a. Data Type Conversion
- b. Missing Values Treatment
- c. Outlier Treatment
- d. Variable Transformation
  - o Data Normalization
  - o Data Standardization
  - o Box Cox Transformation
- e. Variable creation
  - o Dummy Variable Creation
  - o Feature Engineering

### Performance Measurement of Models

- f. Measures for a classification model:
  - o Accuracy
  - o Confusion matrix, TPR, FPR, FNR, TNR
  - o Precision and recall, F1-score
  - o Receiver Operating Characteristic Curve (ROC) curve and AUC
  - o Log-loss
- g. Measures for a Regression model:
  - o R-Squared/Coefficient of determination
  - o Median absolute deviation (MAD)

## 9. Machine Learning

### Supervised Machine Learning

### Unsupervised Learning/Clustering

- K Means Clustering
- Hierarchical Clustering
- DBSCAN (Density Based Clustering)
- Evaluation Metrics for Clustering

### Reinforcement Learning

## Linear Regression

- Intuition of linear regression
- Mathematical formulation
- How to use linear regression in real world
- Interpret the results of linear regression

## Logistic Regression:

- Intuition of Logistic Regression
- Creating a Sigmoid Function from Linear Equation
- Probabilistic Interpretation

## Decision Trees

- Intuition of decision tree
- Various ways in building decision tree
  - Entropy
  - Information Gain
  - Gini Impurity
- Preprocessing for Decision Tree
- Overfitting and Under fitting
- Prediction Using Decision Tress

## K-Nearest Neighbors

- Intuition of KNN
- Distance measures
- How to measure the effectiveness of k-NN?
- Prediction Using KNN

## Support Vector Machines

- Intuition of KNN
- Maximal-Margin Classifier and Its Calculation
- Real World Problems with Margin Classifier
- Different Type of Kernels
- How to Learn a SVM Model?
- Preparing Data for SVM
- Prediction Using KNN

## Ensemble Models

- Bootstrap Method
- Bagging
- Random Forest
- Variable Importance
- Preparing Data

## Re-sampling Techniques

- Leave one out cross validation (**LOOCV**)
- **K-fold**
- Repeated Hold-out Data
- Stratified k-fold cross validation

## Solving Optimization Problems

- What is Optimization
- Why Optimization
- Applications of Optimization
- What is Gradient Descent?
- Varieties in Gradient Descent algorithm
- Implementation of Gradient Descent

## Hyper Parameter Tuning

- Why Tuning
- Manual Tuning
- Grid search

## Model Deployment

- Saving Model in a Pickle File
- Model load from Pickle file and Prediction

## 10. Feature Engineering

- Univariate Selection.
- Recursive Feature Elimination.
- Principal Component Analysis (PCA)
- Feature Importance.

## 11. Deep Learning

- How Biological Neurons work?
- Introduction to Deep Learning
- What is TensorFlow
- What is Keras
- Setting up new environment for Deep Learning
- Perceptron Learning
- Multi-Layered Perceptron (MLP)
- Backpropagation
- Activation functions
- Neural Network Application and Parameter Tuning

### **Artificial Neural Networks(ANN)**

### **Convolutional Neural Networks(CNN)**

- How CNN works
- Steps in CNN
- Convolution in CNN
- Convolutional Operation
- Relu Layers'
- What is pooling vs Flattening
- Full connection
- Softmax vs Cross Entropy
- Max Pooling in CNN
- Flattening in CNN

### **Recurrent Neural Network(RNN)**

- Exploding Gradient
- Vanishing Gradient
- LSTM

### **Time series Analysis**

- Describe Time series data
- Format your time series data
- List the different components of Time series data

- Discuss different kind of Time Series scenarios
- Choose the model according to the time series scenario
- Implement the model for forecasting
- Explain working and implementation of ARMA Model
- Illustrate the working and implementation of different ETS Models
- Forecast the data using the respective model
- What is Time series data?
- Time series variable
- Different components of Time Series data
- Visualize the data to identify time series component
- Implement ARIMA Model for forecasting
- Exponent smoothing model
- Identifying different time series scenario based on which different exponential smoothing model can be applied
- Implement respective model for forecasting
- Visualizing and formatting time series data
- Plotting decomposed time series data plot
- Applying ARIMA and ETS Model for time Series forecasting
- Forecasting for given Time Period
- Case study

## 12. Natural Language Processing/Text mining

- Tokenization
- Part of Speech Tagging
- Chunking
- Stemming
- Important Tasks in NLP
- Important Libraries for NLP
- Traditional NLP/ NLP with Deep learning
- Understand & Implement word2vec
- Understand & build a project on GloVe
- Project on Advanced Sentiment Analysis using Deep Learning

### **Sentiment analysis using NLP**

### **Computer Vision**

## Tools

13. Tensor Flow

14. Hadoop

15. Keras

16. Spark

## III. Case Studies

## IV. Projects (3 -15)

## V. Monthly Tests

## VI. Mock interviews

## VII. HR Activities

## VIII. Certification

■ ■ ■