Data Science course content

Class Room: Training Fee & Duration : 20K & 45 Hours	**Online** Training Fee & Duration : 25K & 45 Hours

Introduction about Data Science:

- What is data science?
- Need of data science?
- Use cases of Data science
- How is data science different from business intelligence?
- Who are data scientists?
- What are the skills required and life cycle of data science?

1. **R programming**
2. **Model building**

R Programming:

Section 1: Data science with R

- Application of machine learning
- Understand Business Analytics and R
- Knowledge on the R language
- Community and ecosystem
- Understand the use of 'R' in the industry
- Compare R with other software in analytics
- Install R and the packages useful for the course
- Perform basic operations in R using command line
- Learn the use of IDE R Studio and Various GUI
- Use the 'R help' feature in R
- Knowledge about the worldwide R community collaboration

Section 2: Introduction to R Programming

- The various kinds of data types in R and its appropriate uses
- The built-in functions in R like: seq(), cbind (), rbind(), merge()
- Knowledge on the various Sub setting methods
- Summarize data by using functions like: str(), class(), length(), nrow(), ncol()
- Use of functions like head(), tail(), for inspecting data
- Indulge in a class activity to summarize data
- If Else

- Nested If Else
- For Loop
- While Loop

Section 3: Data Manipulation in R

- The various steps involved in Data Cleaning
- Functions used in Data Inspection
- Tackling the problems faced during Data Cleaning
- Uses of the functions like grepl(), grep(), sub()
- Coerce the data
- Uses of the apply() functions

Section 4: Data Import Techniques in R

- Import data from spreadsheets and text files into R
- Import data from other statistical formats like sas7bdat and spss
- Packages installation used for database import
- Connect to RDBMS from R using ODBC and basic SQL queries in R
- Basics of Web Scraping

Section 5: Exploratory Data Analysis

- Understanding the Exploratory Data Analysis(EDA)
- Implementation of EDA on various datasets
- Boxplots
- Understanding the cor() in R
- EDA functions like summarize(), llist()
- Multiple packages in R for data analysis
- The Fancy plots like Segment plot
- HC plot in R

Section 6: Data Visualization in R

- Understanding on Data Visualization
- Graphical functions present in R
- Plot various graphs like tableplot, histogram, boxplot
- Customizing Graphical Parameters to improvise the plots
- ggplot2

Model building:

Section 7: Data Pre-processing

- Get the dataset
- Importing the Libraries
- Missing Data
- Categorical Data
- Splitting the Dataset into the Training set and Test set

- Feature Scaling
- Data Pre-processing Template!

Supervised Techniques:

Section 8: Regression

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Support Vector Regression (SVR)
- Decision Tree Regression
- Random Forest Regression
- Evaluating Regression Models Performance
    - R-Squared Intuition
    - Adjusted R-Squared Intuition
    - Interpreting Linear Regression Coefficients

Supervised Techniques Classification:

Section 9: Classification

- Logistic Regression
- K-Nearest Neighbours (K-NN)
- Support Vector Machine (SVM)
- Naive Bayes
- Decision Tree Classification
- Random Forest Classification
- Evaluating Classification Models Performance
    - False Positives & False Negatives
    - Confusion Matrix
    - Accuracy Paradox
    - CAP Curve
    - CAP Curve Analysis

Unsupervised Techniques:

Section 10: Clustering

- K -Means Clustering

Section 11: Association Rule Learning

- Apriori (Market basket Analysis)

Section 12: Text mining

- Sentiment analysis (Twitter)
- Natural Language processing (NLP)

Section 13: Deep Learning

- What is Deep Learning?
- Artificial Neural networks (ANN)
    - The Neuron
    - The Activation Function
    - How do Neural Networks work?
    - How do Neural Networks learn?
    - Gradient Descent
    - Stochastic Gradient Descent
    - Back propagation
- Convolutional Neural Networks (CNN, **Image recognition**)
    - What are convolutional neural networks?
    - Step 1 - Convolution Operation
    - Step 1(b) - ReLU Layer
    - Step 2 - Pooling
    - Step 3 - Flattening
    - Step 4 - Full Connection
    - SoftMax & Cross-Entropy

Section 14: Model Selection & Boosting

- Model Selection
- k-Fold Cross Validation
- Grid Search
- XGBoost

Section 15: Projects

- 2 Real time projects

Section 16: Statistics

- Statistics will be covered during the course where ever it's required.

Tableau:

Section 1: Introduction about Tableau

Section 2: Tableau Basics: Your First Bar chart

Section 3: Timeseries, Aggregation, and Filters

Section 4: Maps, Scatterplots, and Dashboards

Section 5: Joining and Blending Data, PLUS: Dual Axis Charts

Section 6: Table Calculations, Advanced Dashboards, Storytelling

Section 7: Advanced Data Preparation

SQL:

1. Introduction about sql server

2. Introduction TSQL (transact structured query language)

- Types Of TSQL Commands
- Data Definition Language (DDL)
- Data Manipulation Language (DML)
- Data Query Language (DQL)
- Data Control Language (DCL)
- Transaction Control Language (TCL)

3. Database

- Creating Database
- Altering Database
- Deleting Database

4. DML Commands

- Insert
- Identity
- Creating A Table From Another Table
- Inserting Rows From One Table To Another
- Update
- Computed Columns
- Delete
- Truncate
- Differences Between Delete and Truncate

5. DQL

- Select
- Where clause
- Order By Clause
- Distinct Keyword
- Isnull() function
- Column aliases
- Predicates
  - Between … And

In
Like
Is Null

6. Joins

- Inner
- Left
- Right
- Full outer

7. Functions
8. Stored procedures & views