



Pimple Saudagar

Nal Stop

Kharadi

www.etlhive.com

mail@etlhive.com

+91-8237726443

Data Science

Course Details

Introduction to Big Data in the Context of Data Science

- What is Data Analytics?
- Types of Data Sets and Data Models
- Understanding of Business Analytics
- Need of Business Analytics
- Types of Business Analytics
- Descriptive Analytics
- Predictive Analytics
- Prescriptive Analytics
- Supply Chain Analytics
- Health Care Analytics
- Marketing Analytics
- Human Resource Analytics
- Data Management and Business Analytics
- Web Analytics and Business Intelligence
- Data Science as a Strategic Asset
- Data Warehousing and OLAP
- Data Visualization using R and Excel
- Data Visualization using Tableau
- BigData and Data Science

Fundamentals of R

- Understanding R
- Which Companies Use R?
- Understanding Comprehensive R Archive Network (CRAN)
- How to Install R on Operating Systems?
- How to Install R on Windows from CRAN Website?
- IDEs for R
- R Packages: Installation and Practice
- Understanding R Programming
- Studying Operators in R
- Operators: Arithmetic, Relational, Logical, Assignments
- Statements in R Programming
- Conditional Statements in R
- Break and Next Statement
- If else () Function
- Switch Function
- Scan () Function
- Loops in R
- How to Run an R Script and Batch Script?
- R Functions: Commonly Used and String Functions

R Data Structures

- Defining Data Structures in R
- Types of Data Structures
- Vectors and Scalars
- Colon Operator
- Matrices
- Elements: Vector, Matrix, Array
- Understanding Data Frames
- Factors and Lists
- How to Import Files in R?
- How to Import an Excel File?
- How to Import Minitab File?
- Importing Table and CSV Files
- Importing Data from SQL Databases

- **How to Export Files from R?**
- **Types of Apply Functions**
- **Apply () Function: Lapply, Sapply, Tapply**
- **Vapply () Function, Mapply () Function**
- **Understanding Dplyr Package**

Supervised Learning Techniques and the Implementation of Algorithms

- **R Data Structures – Vectors, Factors, Lists, Data Frames, Matrixes and Arrays**
- **Managing Data with R**
- **Saving and Loading R Data Structures**
- **Importing and Saving Data from CSV Files**
- **Importing Data from SQL Databases**
- **Exploring the Structure of Data**
- **Exploring and Understanding Data**
- **Exploring Numeric Variables**
- **Understanding Types of Data**
- **Qualitative and Quantitative Analysis**
- **Studying Descriptive Statistics**
- **Exploring Numeric Variables**
- **Measuring the Central Tendency – The Model**
- **Measuring Spread – Variance and Standard Deviation**
- **Visualizing Numeric Variables – Boxplots and Histograms**
- **Understanding Numeric Data – Uniform and Normal Distributions**
- **Measuring the Central Tendency – The Mode**
- **Exploring Relationships between Variables**
- **Visualizing Relationships – Scatterplots**
- **Nominal and Ordinal Measurement**
- **Interval and Ratio Measurement**
- **Statistical Investigation**
- **Inferential Statistics**
- **Probability and Central Limit Theorem**
- **Exploratory Data Analysis**
- **Normal Distribution**
- **Distance Measures**
- **Euclidean & Manhattan Distance**

- **Minkowski & Mahalanobis**
- **Cosine**
- **Correlation**
- **PPMC (Pearson Product Moment Coorelation)**

Hypothesis Testing

- **Importance of Hypothesis Testing in Business**
- **Null and Alternate Hypothesis**
- **Understanding Types of Errors**
- **Contingency Table and Decision Making**
- **Confidence Coefficient**
- **Upper Tail Test and Test Statistics**
- **Understanding Parametric Tests**
- **Z-Test and Z-Test in R**
- **Chi-Square Test**
- **Degree of Freedom**
- **One-Way ANOVA Test**
- **F-Distribution, F-Ration Test**

Introduction to Data Science and Machine Learning

- **Implementing Machine Learning Algorithms on larger Data Sets with Apache Mahout**
- **How do Machines Learn?**
- **Abstraction and Knowledge Representation**
- **Generalization**
- **Assessing the Success of Learning**
- **Steps to apply Machine Learning to your Data**
- **Choosing a Machine Learning Algorithm**
- **Thinking about the Input Data**
- **Thinking about Types of Machine Learning Algorithms**
- **Matching your Data to an Appropriate Algorithm**

Machine Learning Techniques using R

- **Data Preparation for Modelling**
- **How do Machines Learn?**
- **Choosing a Machine Learning Algorithm**

- **Machine Learning Techniques Using R**
- **Machine Learning: Tasks, Features, Models, and Design**
- **Machine Learning Common Use Case**
- **Supervised and Unsupervised Learning Techniques**
- **Clustering**
- **Similarity Metrics**
- **Distance Measure Types: Euclidean**
- **Cosine Measures**
- **Creating predictive Models**
- **Classification using Nearest Neighbors**

Supervised Learning Techniques and the Implementation of Various Algorithms

- **Supervised Learning Techniques and Algorithms**
- **Understanding Process Flow of Supervised Learning Techniques**
- **K-NN, Naïve Bayes, Support Vector Machines**
- **Defining Classification**
- **Understanding Classification and Prediction**
- **Decision Tree Classifier**
- **How to Build Decision Trees?**
- **Basic Algorithm for a Decision Tree**
- **Decision Trees and Data Mining**
- **Random Forest Classifier**
- **Features of Random Forests**
- **Out of Box Error Estimate and Variable Importance**
- **Naïve Bayes Classifier Model**
- **Bayesian Theorem**
- **Advantages and Disadvantages of Naïve Bayes Classifier Model**
- **Understanding Support Vector Machines**
- **What is Geometric Margin SVMs?**
- **Understanding Linear SVMs**

Unsupervised Machine Learning Techniques – Implementation of Different Algorithms

- **Studying Clustering**
- **Clustering and Classification**
- **Understanding K-means Clustering**
- **K-means and Pseudo Code**
- **K-means Clustering using R**

- **TF-IDF and Cosine Similarity**
- **Application to Vector Space Model**
- **What is Hierarchical Clustering?**
- **Hierarchical Clustering Algorithm**
- **Understanding Agglomerative Clustering Process**
- **DBSCAN Clustering**
- **What is Association Rule Mining?**
- **Association Rule Strength Measures**
- **Checking Apriori Algorithms**
- **Ordering Items**
- **Understanding Candidate Generation**
- **Performing Visualization on Associated Rules**

Regression Methods for Forecasting Numeric Data

- **What is Regression?**
- **Model Selection**
- **Generalized Regression**
- **Simple Linear Regression**
- **Multiple Linear Regression**
- **Correlations**
- **Correlation between X and Y**
- **Ridge and Regularized Regression**
- **LASSO**
- **Time Series**
- **Prediction: Time Dependent/Variant Data**
- **Ordinary Least Square Regression Model**
- **Dummy Variable Regression Model**
- **Interaction Regression Model**
- **Non-Linear Regression Model**
- **Perform Regression Analysis with Multiple Variables**
- **Non-Linear and Linear Models**

Deep Learning – Neural Networks and Support Vector Machines

- **Understanding Neural Networks**
- **From Biological to Artificial Neurons**
- **Activation Functions**

- Network Topology
- Neural Networks: Master Feed-Forward
- Recurrent and Gaussian Neural Network
- The Number of Layers
- The Direction of Information Travel
- The Number of Nodes in Each Layer
- Training Neural Networks with Backpropagation
- Support Vector Machines
- Classification with Hyperplanes
- Finding the Maximum Margin
- The Case of Linearly Separable Data
- The Case of Non-Linearly Separable Data
- Retrieve Data using SQL Statements
- Using Kernels for Non-Linear Spaces
- Performing OCR with SVMs

Python for Data Science

Python Setup

- Python Environment Setup and Essentials
- Anaconda Python Distribution – Windows, Mac OS, Linux
- Jupyter Notebook Installation
- Variable Assignment
- Understanding Data Types: Integer, Float, String, None, Boolean, Typecasting
- Tuples: Create, Access, and Slice
- Dicts: Create, View, Access, and Modify
- Studying Basic Operations: 'in', '+', '*'

Computing with Python – NumPy and SciPy

- Mathematical Computing with Python - NumPy
- Understanding NumPy
- ndarray: Purpose, Properties, Types
- ndarray: Class and Attributes
- How to Access Array Elements?

- Indexing, Slicing, Iteration, Indexing with Boolean Arrays
- Studying Universal Functions
- What is Shape Manipulation?
- Linear Algebra
- Scientific Computing with Python – SciPy
- Understanding SciPy
- Studying SciPy Sub-packages
- Sub-Packages: Integration and Optimize
- Sub-Packages: Statistics, Weave, I O
- Linear Algebra

Data Manipulation and Machine Learning with Python

- Data Manipulation and Machine Learning with Python
- Data Manipulation with Python – Pandas
- Understanding Pandas
- Defining Data Structures
- Data Operations and Data Standardization
- Pandas: File Read and Write Support
- SQL Operation
- Machine Learning with Python – Scikit
- Supervised Learning Models: Linear and Logistic Regression
- Unsupervised Learning Models: Clustering and Dimensionality Reduction
- Model Persistence and Model Evaluation
- Natural Language Processing with Scikit
- NLP Environment Setup & Applications
- NLP Sentence Analysis & Libraries
- Scikit – Built-in Modules & Feature Extraction
- Scikit – Grid Search & Parameters

Data Visualization and Web Scraping

- Data Visualization and Matplotlib
- Python Libraries
- Features of Matplotlib
- Line Properties Plot with (x, y)
- Set Axis, Labels, and Legend Properties
- Alpha and Annotation
- Multiple Plots and SubPlots

- Python Web Scraping and Data Science
- The Parser
- Searching & Modifying the Tree
- Printing, Formatting, Encoding

Python and Hadoop, MapReduce, and Spark

- Importance of Integrating Python with Hadoop
- Understanding BigData Hadoop Architecture
- Working with MapReduce
- Cloudera QuickStart VM Setup
- Studying Apache Spark
- Resilient Distributed Systems (RDD)
- Working with PySpark
- PySpark Integration with Jupyter Notebook

Tableau for Data Science

Tableau and Data Visualization

- What is Data Visualization?
- Scope of Data Visualization
- Tableau Visualization Engine
- Various Visualizations: Text Tables, Pie Charts, Bar, and Line
- Visualizations: Heat Maps, Side by Side Lines, Highlight Tables, Circle Plots
- Visualizations: Tree Maps, Area Charts, Dual Charts, Scatter Plots
- Tableau Workspace
- Dashboard and the Startup Quadrant
- Dashboard Tricks: Reference Lines, Droplines, and Tool Tips

Dashboards and Work Sharing

- Building Interactive Dashboards
- What are Action Filters?
- How to create Story Boards?
- Best Practices to create Dashboards
- Cover Pages

- **Annotations**
- **Tool Tips and keyboard short cuts**
- **Sharing work**
- **Tableau Online**
- **Tableau Reader**

Tableau and Data Connections

- **Understanding Data Connections**
- **How to connect to Tableau Data Server?**
- **Data Connections: Joining and Blending**
- **Defining a Join**
- **Various Kinds of Join**
- **Usage of Join**
- **Right Outer Join**
- **Custom SQL Enabled**
- **Data Blending and Tableau**
- **Usage of Data Blending**
- **Data Blending in Tableau**
- **What is Kerberos Authentication?**
- **Working of Kerberos Authentication**

Data Organization

- **How to Organize and Simplify Data?**
- **What is Filtering?**
- **How to Apply a Filter to a View?**
- **Filtering on Dimensions**
- **Totals and Sub totals**
- **Aggregating Measures and Data Spotlighting**
- **String Functions and Logical Functions**
- **How to Sort Data in Tableau?**
- **Combined Fields**
- **Group and Aliases**
- **Advanced Table Calculations**
- **Calculated Fields and Table Calculations**
- **Quality Assurance for Table Calculations**
- **Hierarchies and Sets**
- **Tableau Bins**

- **Fixed Size and Variable Sized Bins**
- **Drilling and Drilling Methods**
- **Aggregations**
- **Formatting and Annotations**
- **Spatial Analysis and Geo-Coding**
- **Chart Types: Motion Charts, Gantt Charts**
- **Box and Whisker Plots**
- **Mapping and Locations**
- **Pan Zoom Lasso and Radial Selection**

Advanced Data Preparation and Analytics in Tableau

- **Studying Retail Sector Forecasting**
- **Building and Analyzing Box Plots**
- **How to work with Large Data Sources in Tableau?**
- **Understanding and Implementing Pivot and Split**
- **Real World Retail and its Data**
- **Data Source Filters**
- **Trendlines**
- **Advanced Timeseries Blending**
- **Calculating Sales Per Capita**
- **Forecasting in Tableau**
- **How to Present a Storyline?**
- **Creating Animations in Tableau**
- **Real World Case Study: World Health Trends Investigation**
- **Building Visualization and Adding Animation**
- **Manually Sorting Blended Data**
- **Finalizing the Dashboard and Animations in Tableau**