Syllabus for

Hadoop Analytics using R (For Data Scientist)

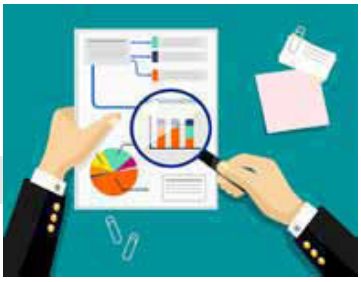
MiND Scripts® Success Simplified

Course Duration For Haddop Analytics using R (For DataScientist) Course :

- 8 Weekend (Weekend batches)

Objective For Hadoop Analytics using R (For DataScientist) Course :

- Provide Insights About the Roles of a Data Scientist
- Provide an understanding of the structure of datasets and databases, including "big data"
- Provide Insights About the Roles of a Data Scientist
- Ability to Analyze Big Data
- Make predictions using machine learning
- Learn to apply hypotheses and data into actionable predictions



Eligibility For Hadoop Analytics using R (For DataScientist) Course :

BSc, BCS, BCA, BE, B.Tech, MSc, MCS, MCA, M.Tech
A background of 1 year in statistics will be helpful

Programming in Hadoop Analytics using R (For Data Scientist)

Analytics for Beginners

- Introduction to Big data Business Analytics
- Applications of Analtyics
- Analtyics Technology and Resources
- Models and Algorithm
- Key roles for successful analytic project
- Main phases of life cycle
- State of the practice in analytics role of data scientists
- Developing core deliverables for stakeholders

Business Statistics

- Descriptive Statistics
- Probabilty and Sampling
- Inferential Statistics
- Hypothesis Testing
- Advanced Hypothesis Testing

Predictive Analytics

- Predictive modeling and Analysis - Regression Analysis
- Multicollinearity
- Correlation analysis
- Multiple correlation
- Least square

An Introduction to R

- Analytics Tools and Exploring R
- Data Structures in R
- Data Manipulation in R
- Dataframe factor

Syllabus for

Hadoop Analytics using R (For DataScientist)

MiND Scripts® Success Simplified

Functions & plots In R

- Measuring the central tendency – the model
- Measuring spread – variance and standard deviation
- Visualizing numeric variables – boxplots
- Visualizing numeric variables – histograms
- Visualizing numeric variables – qqplot
- Understanding numeric data – uniform and normal distributions
- Measuring the central tendency – the model
- Exploring relationships between variables
- Visualizing relationships – scatterplots
- Exploring numeric variables

Read and Write Operations in R

- Reading from CSV
- Reading from URL
- Reading from Excel
- Writing to CSV & PMML

Integrating R

- Implementing Association rule mining in R
- Integrating R with Hadoop using RHadoop and RMR package
- Writing MapReduce Jobs in R and executing them on Hadoop
- Implementing Machine Learning Algorithms on larger Data Sets with Apache Mahout

Databases and Introduction to Machine Learning Concept

- Use SQL databases to store and organize data
- Access stored data with MySQL querying language
- Introduction to Machine Learning
- Supervised and Unsupervised Learning Techniques

Regression Methods and Supervised Learning Techniques

- Creating predictive models
- Classification Using Nearest Neighbors
- Linear Regression
- Multiple linear regression model
- Logistic Regression
- Decision Tree Classifier
- Clustering
- What is Random Forests?
- Features of Random Forest
- Out of Box Error Estimate
- Naive Bayes Classifier

Unsupervised Machine Learning Techniques

- Introduction of K-Means Clustering
- K-means in Euclidean space
- K-means as optimization
- Understanding TF-IDF and Cosine
- Similarity and their application to Vector Space Model

Deep learning

- Deep Networks
- Optimization for Training Deep Models
- Convolutional Networks
- Understanding Support Vector Machines
- Retrieve data using sql statements
- Using kernels for non-linear spaces

Project

Project name: Live Project

Project description: Student will be assigned a project which they will have to execute under the careful guidance of the faculty.