SLOG SOLUTIONS PRIVATE LIMITED
TECHNOLOGY: DATA SCIENCE
DURATION: 6 WEEKS

- **Introduction to Data Science**
  - Introduction to Big Data
  - Roles played by a Data Scientist
  - Analyzing Big Data using Hadoop and R
  - Methodologies used for analysis
  - The Architecture and Methodologies used to solve the Big Data problems

- **Basic Data Manipulation using R**
  - Understanding vectors in R
  - Reading Data, Combining Data
  - Subsetting data
  - Sorting data and some basic data generation functions

- **Machine Learning Techniques Using R Part-1**
  - Machine Learning Overview,
  - ML Common Use Cases
  - Understanding Supervised and Unsupervised Learning
  - Techniques, Clustering
  - Similarity Metrics
  - Distance Measure Types: Euclidean, Cosine Measures, Creating predictive models

- **Machine Learning Techniques Using R Part-2**
  - Understanding K-Means Clustering
  - Understanding TF-IDF and Cosine Similarity and their application to Vector Space Model
  - Implementing Association rule mining in R

- **Machine Learning Techniques Using R Part-3**
  - Understanding Process flow of Supervised Learning Techniques
  - Decision Tree Classifier
  - How to build Decision trees
  - Random Forest Classifier
  - What is Random Forests
  - Features of Random Forest
  - Out of Box Error Estimate and Variable Importance
  - Naive Bayes Classifier

- **Introduction to Hadoop Architecture**
  - Hadoop Architecture
  - Common Hadoop commands
  - MapReduce and Data loading techniques (Directly in R and in Hadoop using SQOOP, FLUME, and other Data Loading Techniques)
  - Removing anomalies from the data

- **Integrating R with Hadoop**
  - Integrating R with Hadoop using RHadoop and RMR package
  - Exploring RHIPE (R Hadoop Integrated Programming Environment)
  - Writing MapReduce Jobs in R and executing them on Hadoop

- **Mahout Introduction and Algorithm Implementation**
  - Implementing Machine Learning Algorithms on larger Data Sets with Apache Mahout

- **Additional Mahout Algorithms and Parallel Processing using R**
  - Implementation of different Mahout algorithms
  - Random Forest Classifier with parallel processing Library in R

SLOG SOLUTIONS PVT. LTD.

SLOG SOLUTIONS PVT.LTD.
HELPLINE 7456000240/7456000241
www.slogsolutions.com