

CETPA INFOTECH PVT. LTD.
CURRICULUM OF DATA SCIENCES

<p>1. Introduction to Data Science</p> <ul style="list-style-type: none"> • 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 <p>2. Basic Data Manipulation using R</p> <ul style="list-style-type: none"> • Understanding vectors in R • Reading Data, Combining Data • Subsetting data • Sorting data and some basic data generation functions <p>3. Machine Learning Techniques Using R Part-1</p> <ul style="list-style-type: none"> • 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 <p>4. Machine Learning Techniques Using R Part-2</p> <ul style="list-style-type: none"> • Understanding K-Means Clustering • Understanding TF-IDF and Cosine Similarity and their application to Vector Space Model • Implementing Association rule mining in R <p>5. Machine Learning Techniques Using R Part-3</p> <ul style="list-style-type: none"> • Understanding Process flow of Supervised Learning Techniques • Decision Tree Classifier 	<ul style="list-style-type: none"> • 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 <p>6. Introduction to Hadoop Architecture</p> <ul style="list-style-type: none"> • 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 <p>7. Integrating R with Hadoop</p> <ul style="list-style-type: none"> • 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 <p>8. Mahout Introduction and Algorithm Implementation</p> <ul style="list-style-type: none"> • Implementing Machine Learning Algorithms on larger Data Sets with Apache Mahout <p>9. Additional Mahout Algorithms and Parallel Processing using R</p> <ul style="list-style-type: none"> • Implementation of different Mahout algorithms • Random Forest Classifier with parallel processing Library in R
--	---

HEAD OFFICE: 200 Purwavali , 2nd Floor, (Opp. Railway Ticket Agency), Railway Road , Ganeshpur, Roorkee – 247667, Ph.No.: 09219602769, 01332-270218 Fax - 1332 – 274960

CORPORATE OFFICE: D-58, Sector-2, Near Red FM. Noida -201301, Uttar Pradesh
 Contact Us: +91-9212172602 , 0120-4535353

BRANCH OFFICE: 401 A, 4th Floor, Lekhraj Khazana, Faizabad Road, Indira Nagar, Lucknow-220616 (U.P.) Ph. No: +91-522-4233162, +91-9258017974

BRANCH OFFICE: 105, Mohit Vihar, Near Kamla Palace, GMS Road, Dehradun-248001, UK
 Contact: +91-9219602771, 0135-6006070
 Toll Free- 1800-8333-999 (from any network)

