Title of the Course: Certificate Course in Big Data Analytics

Aims/Objectives of the Course:

1. To understand the Concept of Big Data Analytics
2. To learn basic concepts of R-Programming
3. Apply the R-Programming for Big Data Analytics

Duration of Course: 1Year

Fee Structure:

Course Structure:

1. Paper-I R-Programming
2. Paper-II Introduction to Big Data Analytics
3. Paper-III Practical Course on Big Data Analytics using R-Programming

Eligibility for Admission: 12th

Structure of Course:

Sr. No	Papers	Credit	Hours for Credit	Total	Total Marks
1.	Paper I R-Programming	6	15	90	100
2.	Paper-II Introduction to Big Data Analytics	6	15	90	100
3.	Paper-III Practical Course on Big Data Analytics using R-Programming	8	15	90	100

Minimum Staff: 02

Mode of Examination:

Detail Syllabus: attached

Head
Department of Computer Science
Dr.Annasaheb G.D.Bendale
Mahila Mahavidyalaya,Jalgaon

Certificate Course in Big Data Analytics

Paper I

R- Programming

Objectives:

The objectives of this course are-

1. To understand the basic concepts of R- Programming
2. To apply R- Programming for Big Data Analytics

Outcomes:

1. Students will able to install R-Studio, and code in R programming
2. Student will able to use of R- Programming for Big Data analytics

Unit 1: History and Overview of R (15 hours)

- 1.1 History of R, Introduction to R, Limitations and features of R,
- 1.2 How to Run R
  - 1.2.1 Interactive Mode
  - 1.2.2 Batch Mode
- 1.3 Introduction to Functions
  - 1.3.1 Variable Scope
  - 1.3.2 Default Arguments
- 1.4 Preview of Some Important R Data Structures
  - 1.4.1 Vectors, the R Workhorse
  - 1.4.2 Character Strings
  - 1.4.3 Matrices
  - 1.4.4 Lists
  - 1.4.5 Data Frames
  - 1.4.6 Classes

Unit 2: VECTORS (10 hours)

- 2.1 Scalars, Vectors, Arrays, and Matrices
  - 2.1.1 Adding and Deleting Vector Elements
  - 2.1.2 Obtaining the Length of a Vector
  - 2.1.3 Matrices and Arrays as Vectors
- 2.2 Declarations
- 2.3 Recycling
- 2.4 Common Vector Operations
  - 2.4.1 Vector Arithmetic and Logical Operations
  - 2.4.2 Vector Indexing
  - 2.4.3 Generating Useful Vectors with the : Operator
  - 2.4.4 Generating Vector Sequences with seq()
  - 2.4.5 Repeating Vector Constants with rep()
- 2.5 Using all() and any()
  - 2.5.1 Extended Example: Finding Runs of Consecutive Ones
  - 2.5.2 Extended Example: Predicting Discrete-Valued Time Series
- 2.6 Vectorized Operations
  - 2.6.1 Vector In, Vector Out

2.6.2 Vector In, Matrix Out
2.7 NA and NULL Values
2.7.1 Using NA
2.7.2 Using NULL

Unit 3: MATRICES AND ARRAYS (15 hours)

3.1 Creating Matrices
3.2 General Matrix Operations
3.2.1 Performing Linear Algebra Operations on Matrices
3.2.2 Matrix Indexing
3.2.3 Extended Example: Image Manipulation
3.2.4 Filtering on Matrices
3.2.5 Extended Example: Generating a Covariance Matrix
3.3 Applying Functions to Matrix Rows and Columns
3.3.1 Using the apply() Function
3.3.2 Extended Example: Finding Outliers
3.4 Adding and Deleting Matrix Rows and Columns
3.4.1 Changing the Size of a Matrix
3.4.2 Extended Example: Finding the Closest Pair of Vertices in a Graph
3.5 More on the Vector/Matrix Distinction

Unit 4. LISTS (10 hours)

4.1 Creating Lists
4.2 General List Operations
4.2.1 List Indexing
4.2.2 Adding and Deleting List Elements
4.2.3 Getting the Size of a List
4.2.4 Extended Example: Text Concordance
4.3 Accessing List Components and Values
4.4 Applying Functions to Lists
4.4.1 Using the lapply() and sapply() Functions
4.4.2 Extended Example: Text Concordance, Continued
4.4.3 Extended Example: Back to the Abalone Data
4.5 Recursive Lists

Unit 5: DATA FRAMES (10 hours)

5.1 Creating Data Frames
5.1.1 Accessing Data Frames
5.1.2 Extended Example: Regression Analysis of Exam Grades
5.2 Other Matrix-Like Operations
5.2.1 Extracting Subdata Frames
5.2.2 More on Treatment of NA Values
5.2.3 Using the rbind() and cbind() Functions and Alternatives
5.2.4 Applying apply()
5.2.5 Extended Example: A Salary Study
5.3 Merging Data Frames
5.3.1 Extended Example: An Employee Database
5.4 Applying Functions to Data Frames
5.4.1 Using lapply() and sapply() on Data Frames
5.4.2 Extended Example: Applying Logistic Regression Models
5.4.3 Extended Example: Aids for Learning Chinese Dialects

Unit 6: R – OPERATORS (10 hours)

- 6.1 Arithmetic Operators
- 6.2 Relational Operators
- 6.3 Logical Operators
- 6.4 Assignment Operators

Unit 7: Control Structures (10 hours)

- 7.1 if , if..else, if…else…if
- 7.2 Switch Statement
- 7.3 for loop
- 7.4 Nested Loops
- 7.5 While loop
- 7.6 Repeat loop
- 7.7 Next,break

Unit 8: Functions (10 hours)

- 8.1 Function creation
- 8.2 Using return
- 8.3 Arguments
- 8.4 Specialized functions

References: -

1. "The book of R A first course in Programming and Statistics",by Tilman M. Davies, ISBN-13:978-1-59327-651-5
2. "R Programming for Data Science", Roger D. Peng

Paper II

Introduction to Big Data Analytics

Objectives: -

1. To understand the fundamental concepts of big data analytics.
2. To learn to analyze the big data using intelligent techniques.
3. To learn to use various techniques for mining data

Outcomes:

Student will be able to

1. Understand the architecture of Big data
2. Student will be able to understand business applications of Big Data
3. Apply clustering and association rule mining for Big data Analysis

Unit 1: Introduction to Big Data Analytics **15**

1.1 Big Data Overview
1.2 State of the Practice in Analytics
1.3 Key Roles for the New Big Data Ecosystem
1.4 Examples of Big Data Analytics

Unit 2: Data Analytics Lifecycle **20**

2.1 Data Analytics Lifecycle Overview
2.2 Phase 1: Discovery
2.3 Phase 2: Data Preparation
2.4 Phase 3: Model Planning
2.5 Phase 4: Model Building
2.6 Phase 5: Communicate Results
2.7 Phase 6: Operationalize
2.8 Case Study: Global Innovation Network and Analysis (GINA)

Unit 3: Review of Basic Data Analytic Methods **20**

3.1 Introduction to R
3.2 Exploratory Data Analysis
3.3 Statistical Methods for Evaluation

Unit 4: Advanced Analytical Theory and Methods: Clustering **15**

4.1 Overview of Clustering
4.2 K-means
4.3 Additional Algorithms

Unit 5: Advanced Analytical Theory and Methods: Association Rules **20**

5.1 Overview
5.2 Apriori Algorithm
5.3 Evaluation of Candidate Rules
5.4 Applications of Association Rules
5.5 An Example: Transactions in a Grocery Store
5.6 Validation and Testing
5.7 Diagnostics

Reference book:

1. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data- EMC Education Services, Wiley Publication.

Paper III
Practical Course on Big Data Analytics using R-Programming

List of Practical's on R Programming and Introduction to Big Data Analytics

1. Installation of R Studio
2. Basic syntax of R
3. Demonstration of Vector data objects
4. Demonstration of matrix, array in R.
5. Demonstration and use of data frames in R.
6. Demonstration of sample data creation and data manipulation in R.
7. Demonstration of various if, if..else, if..else..if, switch, for, while repeat, next, break statements in R.
8. Demonstration of functions in R
9. Data Manipulation with dplyr package
10. Data Manipulation with data.table package
11. Write the R program to visualize single variable.
12. Write the R Program to classify the data using K-means clustering.
13. Write the R program to implement Association rule mining.

Head
Department of Computer Science
Dr.Annasaheb G.D.Bendale
Mahila Mahavidyalaya,Jalgaon