Subject: Computer Science   **Semester: V/VI**
Course Title: Big Data Analytics   **Course Code: 20CSSEC11BD3**
                    with R
No. of Hours: 45   **LTP: 300**   **Credits: 3**

Objectives
- To learn the classification of digital data.
- To learn the classification of analytics.
- To analyze the importance and Exploring data in R programming.

Course Outcomes
CO1: Explain data and classification of digital data.
CO2: Explain Big Data Analytics.
CO3: Load data in to R.
CO4: Organize data in the form of R objects and manipulate them as needed.
CO5: Perform analytics using R programming.

UNIT-I   **(9 Hrs.)**
Introduction to Big data: Data, classification Of Digital Data--structured, unstructured, semi-structured data, characteristics of data, evaluation of big data, definition and challenges of big data , what is big data and why to use big data ?, business intelligence Vs big data. - Programming Exercises

UNIT-II   **(9 Hrs.)**
Big data Analytics: What is and isn't big data analytics? Why hype around big data analytics? Classification of analytics, top challenges facing big data, importance of big data analytics, technologies needed to meet challenges of big data - Programming Exercises.

UNIT-III   **(9 Hrs.)**
Introduction to R and getting started with R: What is R? Why R? , advantages of R over other programming languages, Data types in R- logical, numeric, integer, character, double, complex, raw, coercion, ls() command, expressions, variables and functions, control structures, Array, Matrix, Vectors, R packages - Programming Exercise.

UNIT-IV   **(9 Hrs.)**
Exploring data in R: Data frames-data frame access, ordering data frames, R functions for data frames dim(), nrow(), ncol(), str(), summary(), names(),

head(), tail(), edit() .Load data frames—reading from .CSV files, sub setting data frames, reading from tab separated value files, reading from tables - Programming Exercises.

UNIT-V   **(9 Hrs.)**
Data Visualization using R:
Reading and getting data into R (External Data):XML files, Web Data, JSON files, Databases, Excel files.
Working with R Charts and Graphs: Histograms, Bar Charts, Line Graphs, Scatterplots, Pie Charts - Programming Exercises.

Co-Curricular Activities
- Assignments on problem solving
- Group discussions
- Student presentations and seminars
- Online quizzes
- Project work

Prescribed Books
1. Seema Acharya ,SubhashiniChellappan --- Big Data And Analytics second edition, Wiley.
2. Seema Acharya--Data Analytics using R, McGraw Hill education (India) Private Limited.
3. Big Data Analytics, Introduction to Hadoop, Spark, and Machine-Learning, Raj kamal, Preeti Saxena, McGraw Hill, 2018.
4. Big Data, Big Analytics: Emerging Business intelligence and Analytic trends for Today's Business, Michael Minelli, Michelle Chambers, and AmbigaDhiraj, John Wiley & Sons, 2013.

Reference Book
1. An Introduction to R, Notes on R: A Programming Environment forData Analysis and Graphics. W. N. Venables, D.M. Smith and the R Development Core Team.

MARIS STELLA COLLEGE (AUTONOMOUS), VIJAYAWADA – 8
(Affiliated to Krishna University)
Blueprint

Subject: Computer Science   **Semester: V/VI**
Course Title: Big Data Analytics   **Course Code: 20CSSEC11BD3**
                  with R
Time: 3 Hrs.   **Max. Marks: 100**

SECTION – A

Answer **ALL** questions   **20 x 1 = 20M**

Q. No.	UNIT	Marks Weightage	RBT LEVEL
1	I	1	
2	I	1	
3	II	1	
4	II	1	
5	III	1	**No. of questions to be set**
6	III	1	RBT1 – 8
7	IV	1	RBT2 – 8
8	IV	1	RBT3 – 2
9	V	1	RBT4 – 2
10	V	1	
11	I	1	
12	I	1	
13	II	1	
14	II	1	
15	III	1	
16	III	1	
17	IV	1	
18	IV	1	
19	V	1	
20	V	1	

SECTION – B

Answer any **FOUR** questions   **4 x 8 = 32M**

Q. No.	UNIT	Marks Weightage	RBT LEVEL
21	I	8	**No. of questions to be set**
22	II	8	RBT1 – 2
23	III	8	RBT2 – 2
24	IV	8	RBT3 – 1
25	V	8	RBT4 – 1
26	I / II / III / IV / V	8	

SECTION – C

Answer any **FOUR** questions   **4 x 12 = 48M**

Q. No.	UNIT	Marks Weightage	RBT LEVEL
27	I	12	**No. Of questions to be set**
28	II	12	RBT1 – 2
29	III	12	RBT2 – 2
30	IV	12	RBT3 – 1
31	V	12	RBT4 – 1
32	I / II / III / IV / V	12	

Subject: Computer Science   **Semester: V/VI**
Course Title: Big Data Analytics   **Course Code: 20CSSEC11BD3**
                   with R
Time: 3 Hrs.   **Max. Marks: 100**

SECTION – A

Answer **ALL** questions   **20 x 1 = 20M**

1. The total forms of big data is ____
   A. 1
   B. 2
   C. 3
   D. 4

2. Big data analysis does the following except?
   A. Spread data
   B. Analyze data
   C. Organizes data
   D. Collect data

3. What are the main components of Big Data?
   A. Map Reduce
   B. HDFS
   C. YARN
   D. All of these

4. According to analysts, for what can traditional IT systems provide a foundation When they're integrated with big data technologies like Hadoop?
   A. Big data management and data mining
   B. Data warehousing and business intelligence
   C. Management of Hadoop clusters
   D. Collecting and storing unstructured data

5. Which function is used to create the vector with more than one element
   A. Library()
   B. plot()
   C. c()
   D. par()

6. How many types of R objects are present in R data type?
   A. 4
   B. 5
   C. 6
   D. 7

7. What is the function to set row names for a data frame?

A. row.names()

B. colnames()

C. col.names()

D. column name cannot be set for a data frame

8. If the code is stored in the external file, which of the following function is used to call them to a working directory?

A. exec()

B. source()

C. execute()

D. sourcefile()

9. Which of the following finds the position of a quantile in a dataset?

A. quantile()

B. barplot()

C. barchart()

D. rep()

10. What are the common types of data visualization?

A. Charts

B. Tables

C. Infograpics

D. All the above

11. Transaction of data of the bank is a of _____ type.

12. _____ is a collection of data that is used in volume, yet growing exponentially with time.

13. Data in ____ bytes size is called big data

14. _____ function is used to add title to each axis instance in a figure.

15. R allows integration with the procedures written in the _____.

16. Command lines entered at the console are limited to about _____ bytes.

17. Functionality of R is divided into a number of _____

18. _____ object is masked from 'package: stats'.

19. _____ is used to query and edit graphical settings.

20. A data visualization tool that updates in real time and gives multiple outputs is called _____.

SECTION – B

Answer any **FOUR** questions   **4 x 8 = 32M**

21. What is Big Data? Why we need to analyse Big Data?
    List out the characteristics of big data and challenges in handling big data

22. Illustrate the various phases involved in Big Data Analytics with neat diagram.
23. Demonstrate the data types in R with suitable examples
24. Write a R program to extract specific column from a data frame using column name.
25. What do you understand by the term Data Visualization? How is it important in Big Data Analytics?
26. Write a R program to extract specific column from a data frame using column name.

SECTION – C

Answer any **FOUR** questions   **4 x 12 = 48M**

27. What are the three Vs of Big data? Give two examples of big data case studies. Indicate which Vs are satisfied by these case studies.
28. What is Big Data Analytics? Describe the importance of big data and classification of analytics?
29. a) R has five "atomic" classes of objects. What are they? Quote examples
    b) Demonstrate repeat loop and seq_along ( ) function.
30. a) Which matrix operations are applicable to data frames? Explain with examples.
     b) How to run a logistic regression model on data in a data frame? Give illustration.
31. a) Write about scatter plot and histograms with examples? Explain its important.
    b) How to plot multiple curves in same graph?
32. a) Explain different types of operators in R?
    b) Write about control statements in R