

B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2025
 Fifth Semester
Computer and Communication Engineering
20CSEL504 - FOUNDATION OF DATA SCIENCE
 Regulations - 2020

Duration: 3 Hours

Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

	<i>Marks</i>	<i>K- Level</i>	<i>CO</i>
1. Which of the following is the most important language for Data Science? (a) Java (b) Ruby (c) R (d) None of the mentioned	1	K1	CO1
2. Which of the following is NOT typically considered a part of the scope of data science? (a) Data analysis and modelling (b) Data visualization (c) Database administration (d) Machine learning	1	K1	CO1
3. Consider, for instance, that you want to add a red line to a plot, from (-4, -4) to (4, 4), so choose what you could write? (a) plot(x, y) Plines(-4:4, -4:4, lwd = 3, col = "red") (b) plot(x, y) lines(-4:4, -4:4, lwd = 3, col = "red") (c) plot(x, y) plot(-4:4, -4:4, lwd = 3, col = "red") (d) None of the Mentioned	1	K2	CO2
4. Relate, which of the following is a natural choice to produce a plot that contains several series of data and the data are categorical? (a) boxplot() (b) histogram() (c) barplot() (d) scatterplot()	1	K2	CO2
5. Why is it important to map problems to machine learning models-Infer. (a) To save storage space (b) To improve typing speed (c) To choose the right algorithm based (d) To reduce internet usage on data and goals	1	K2	CO3
6. What does internal validation of a clustering model evaluate? (a) Comparison with labeled data (b) Agreement with previous research (c) Quality of clusters using internal metrics (d) Speed of model execution	1	K1	CO3
7. What is the role of the Data Node in a Hadoop cluster? (a) Managing metadata of files stored in HDFS. (b) Storing actual data blocks of files in HDFS. (c) Directing client requests to the appropriate Name Node. (d) Performing Map Reduce tasks on data stored in HDF	1	K1	CO4
8. Which Hadoop component manages the execution of MapReduce tasks across the cluster? (a) NameNode (b) ResourceManager (c) Secondary NameNode (d) DataNode	1	K1	CO4
9. Infer under which scenario would two naive bayes classifiers likely produce different results? (a) When the training data is skewed (b) When the conditional independence assumption holds true (c) When the class probabilities are equal (d) When the features are highly correlated	1	K2	CO5
10. Given a scenario where you need to bind three row vectors into a single matrix in R, show which function would you use? (a) cbind() (b) matrix() (c) rbind() (d) transpose()	1	K2	CO6

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11. Define Data science.	2	K1	CO1
12. What is the role of a data scientist?	2	K1	CO1

13. What is the Matrix Plot?	2	K1	CO2
14. Infer the purpose of Audience Analysis in presentations.	2	K2	CO2
15. What is model evaluation?	2	K1	CO3
16. Justify your answer with the criteria for choosing a model.	2	K2	CO3
17. Infer the purpose of writing a Hadoop Map Reduce Program.	2	K2	CO4
18. What is HDFS?	2	K1	CO4
19. Compare the training data and testing data.	2	K2	CO5
20. State the advantage and uses of Naïve Bayes.	2	K1	CO5
21. Define about any 4 functions used to manipulate objects in R with example	2	K1	CO6
22. Infer on the functions used to create matrices in R.	2	K2	CO6

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23. a) Define and List on the sampling techniques for modeling and validation.	11	K1	CO1
OR			
b) (i) List the various data cleaning methods with examples.	6	K1	CO1
(ii) What is the process of working with data from files?	5	K1	CO1
24. a) Model the process of documentation and deployment in analytics projects.	11	K3	CO2
OR			
b) Apply matrix plots and multiple plots in one window with R examples.	11	K3	CO2
25. a) Develop the mapping problems to machine learning with examples.	11	K3	CO3
OR			
b) Construct the algorithm of K-Means for an example and explain it .	11	K3	CO3
26. a) Build the architecture of Map Reduce and explain it in detail.	11	K3	CO4
OR			
b) Develop matrix-vector multiplication by Map Reduce Algorithm with an example.	11	K3	CO4
27. a) Illustrate the working of linear regression algorithms with an example, also give its advantages and disadvantages.	11	K2	CO5
OR			
b) Explain the Naïve Bayes classifier algorithm with an example, also stating its advantages and disadvantages.	11	K2	CO5
28. a) Model and illustrate the the various data distribution visualization methods in R.	11	K3	CO6
OR			
b) Develop the arrays and matrices with its operations in R with suitable examples.	11	K3	CO6