



Dezyne École College

Bachelor of Computer Application (B.C.A.) Second Year-6th Semester

Data Mining with R

Part A: Short Answer Questions

(10 Questions × 2 Marks = 20 Marks)

(Answer in about 50 words)

1. Define data mining.
2. Mention two common data mining techniques.
3. What is an association rule in data mining?
4. Define FP-tree growth.
5. What is dynamic itemset counting?
6. Mention any two issues in data mining.
7. Define k-Medoid algorithm.
8. What is a decision tree?
9. Mention two splitting criteria used in decision trees.
10. What is web mining?
11. Define temporal data mining.
12. What is a pruning technique?
13. Define categorical clustering.
14. What is the role of ID3 in decision trees?
15. Mention one difference between CART and Random Forest.
16. Define R language.
17. What is a vector in R?
18. Write the function to read a file in R.
19. Define a matrix in R with example.
20. What is the use of loops in R programming?
21. Mention two types of graphics functions in R.
22. What is GGplot2 in R?
23. What is a hypothesis?
24. State two types of hypotheses.
25. Define parametric test.
26. Define non-parametric test.
27. What is data sampling?
28. Write the syntax to create a list in R.
29. What is the difference between attributes and objects in R?
30. State any one control statement in R.

Part B: Long Answer Questions

(Attempt any 5 out of 10 × 10 Marks = 50 Marks)

(Answer in about 400 words)

Unit I – Data Mining Basics

31. Define data mining and explain its significance in today's digital world.
32. Discuss the major data mining techniques with suitable examples.
33. Explain the issues and challenges involved in data mining.
34. Write short notes on: (a) Association Rules, (b) FP-tree growth.
35. Explain the process of dynamic itemset counting and its advantages.

Unit II – Clustering & Decision Trees

36. Explain the k-Medoid algorithm and its use in clustering.
37. Compare hierarchical and categorical clustering with examples.
38. Explain the process of decision tree construction.
39. What are CART and ID3 algorithms? Compare their features.
40. Describe the concept and working of Random Forest with pruning techniques.
41. Explain the applications of web mining, temporal, and spatial data mining.
42. How is data mining used in Neural Networks? Give examples.

Unit III – R Programming & Hypothesis Testing

43. Discuss the basic elements of R: objects, attributes, numbers, vectors, and arrays.
44. Explain data input/output in R with examples.
45. How do you read data from a file in R? Provide syntax and example.
46. Explain control statements and loops in R with examples.
47. What is data visualization? Explain using graphics and GGplot2 in R.
48. Describe different formats of graphics output in R.
49. What is a hypothesis? Explain its types with examples.
50. Differentiate between parametric and non-parametric tests.
51. Explain the concept of confidence level and significance level.
52. Describe the process of hypothesis testing with an example.
53. What are R scripts? Explain their purpose in data analysis.
54. Explain the importance of data sampling in research.
55. Compare arrays, lists, and matrices in R with suitable examples.
56. Discuss real-world applications of data mining using R programming.
57. Explain the importance of data science overview in data mining.
58. Discuss the use of loops and functions in automating tasks in R.
59. Explain hypothesis generation and testing in the context of data science.
60. Describe a full data mining project life cycle using R tools.