

21U276S

(Pages: 2)

Name:

Reg. No:

SECOND SEMESTER B.Voc. DEGREE EXAMINATION, APRIL 2022

(B.Voc. – Information Technology)

CC18U SDC2 DS07 - INTRODUCTION TO DATA SCIENCE

(2018 to 2020 Admissions – Supplementary/Improvement)

Time: Three Hours

Maximum: 80 Marks

PART A

Answer *all* questions. Each question carries 1 mark.

1. _____ is the process of “taking all aspects of life and turning them into data.”
2. _____ is an approach to analyzing data sets to summarize their main characteristics, often with visual methods.
3. The _____ matrix that has a 1 in row i and column j if there is an edge between nodes i and j , and 0 otherwise.
4. The _____ matrix has nonzero entries only on the diagonal.
5. The difference between the degree matrix and the adjacency matrix is the _____ matrix.
6. The smallest eigenvalue for every Laplacian matrix is _____
7. The number of arcs along the path is the _____
8. A _____ in a directed graph is a sequence of nodes v_0, v_1, \dots, v_k such that there are arcs $v_i \rightarrow v_{i+1}$ for all $i = 0, 1, \dots, k-1$.
9. The _____ for a node v is the set of nodes u for which there is a path of length at most d from v to u .
10. The _____ of a directed graph is the smallest integer d such that for every two nodes u and v there is a path of length d or less from u to v .

(10 × 1 = 10 Marks)

PART B

Answer any *eight* questions. Each question carries 2 marks.

11. What is Supervised Learning?
12. What is Linear Regression?
13. What is classification?
14. What is quota sampling?
15. What is statistical modeling?
16. What is random variable?
17. Define wrappers.

18. What is spam?
19. Define betweenness.
20. Define machine learning.
21. Define singular value decomposition.
22. Define filters.

(8 × 2 = 16 Marks)

PART C

Answer any *six* questions. Each question carries 4 marks

23. What are the different data objects in R?
24. What are the different sampling techniques?
25. Explain Exploratory Data Analysis.
26. Explain decision making statements in R.
27. What is feature extraction? Why it is useful?
28. What is K-means? How can you select K for K-means?
29. Explain decision tree algorithm.
30. Explain Clustering of Social-Network Graphs.
31. Differentiate between univariate, bivariate and multivariate analysis with example.

(6 × 4 = 24 Marks)

PART D

Answer any *two* questions. Each question carries 15 marks.

32. What is data science? Explain the data science process.
33. What is social network? What are its characteristics? How social network is represented as graph?
34. What is data visualization? Explain basic principles of data visualization.
35. Explain naïve bayes classification with example.

(2 × 15 = 30 Marks)
