

**20U254**

(Pages: 2)

Name: .....

Reg. No: .....

**SECOND SEMESTER B.Voc. DEGREE EXAMINATION, APRIL 2021**

(Regular/Supplementary/Improvement)

**SDC2 DS07 – INTRODUCTION TO DATA SCIENCE**

(Information Technology)

(2018 Admission onwards)

Time: Three hours

Maximum: 80 Marks

**PART A**

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

1. \_\_\_\_\_ is an approach to analyzing data sets to summarize their main characteristics, often with visual methods.
2. The number of arcs along the path is the \_\_\_\_\_
3. The smallest eigenvalue for every Laplacian matrix is \_\_\_\_\_
4. \_\_\_\_\_ is populating the inbox of any target victim with unsolicited or junk emails.
5. \_\_\_\_\_ technique is also used in product advertisement.
6. \_\_\_\_\_ is the process of finding a model that describes and distinguishes data classes or concepts.
7. \_\_\_\_\_ is a widely used and effective machine learning algorithm based on the idea of bagging.
8. \_\_\_\_\_ is the role of exploratory graphs in data analysis.
9. \_\_\_\_\_ is an example of feature extraction.
10. Euclidean distance measure is \_\_\_\_\_

**(10 × 1 = 10 Marks)**

**PART B**

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

11. What is Linear Regression?
12. What is data science?
13. What is cluster analysis?
14. What is sample?
15. What is probability distribution?
16. What is stratified sampling?
17. Define random forest.
18. What is spam filtering?
19. Define betweenness.

20. What are Eigenvalue and Eigenvector?
21. What is normalized cut?
22. What is big data?

**(8 × 2 = 16 Marks)**

### **PART C**

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

23. What are the different sampling techniques?
24. What is Data Science? Also, list the differences between supervised and unsupervised learning.
25. What are the applications of R?
26. Explain data science process.
27. Explain Exploratory Data Analysis.
28. Explain decision tree algorithm.
29. What are the different types of social networks?
30. Explain principal component analysis.
31. Explain k-Nearest Neighbors (k-NN) with example.

**(6 × 4 = 24 Marks)**

### **PART D**

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

32. What is data visualization? What are the different tools for data visualization? Explain basic principles of data visualization.
33. Explain k-mean algorithm.
34. Explain feature selection. What are the different types of feature selection techniques?
35. What are the different data objects in R? Explain types of operators in R. Explain decision making statements in R.

**(2 × 15 = 30 Marks)**

\*\*\*\*\*