

**19U256**

(Pages: 2)

Name : .....

Reg. No. : .....

**SECOND SEMESTER B.Voc. DEGREE EXAMINATION, APRIL 2020**

(Regular/Supplementary/Improvement)

**CC18U SDC2 DS07 – INTRODUCTION TO DATA SCIENCE**

(B.Voc. – Information Technology)

(2018 Admission onwards)

Time: 3 Hours

Maximum: 80 Marks

**PART A**

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

1. The number of arcs along the path is the \_\_\_\_\_
2. The \_\_\_\_\_ matrix has nonzero entries only on the diagonal.
3. The smallest eigenvalue for every Laplacian matrix is \_\_\_\_\_
4. 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  $u$ .
5. \_\_\_\_\_ is populating the inbox of any target victim with unsolicited or junk emails.
6. \_\_\_\_\_ technique is also used in product advertisement.
7. The process of “taking all aspects of life and turning them into data” is \_\_\_\_\_
8. KDD stands for \_\_\_\_\_
9. \_\_\_\_\_ is the process of finding a model that describes and distinguishes data classes or concepts.
10. The strength of the linear relationship between two numerical variables be measured by \_\_\_\_\_

**(10 x 1 = 10 Marks)**

**PART B**

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

11. What is Linear Regression?
12. What is probability distribution?
13. What is datafication?
14. Difference between normal & binomial distribution?
15. What are Eigenvalue and Eigenvector?
16. Define feature selection.
17. What is discrete and continuous random variable?

18. Write a few applications of R.
19. What are the different types of social networks?
20. What are the different tools for data visualization?
21. What is normalized cut?
22. Define wrappers.

**(8 x 2 = 16 Marks)**

### **PART C**

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

23. What are the different data objects in R?
24. Explain data science process.
25. Explain Exploratory Data Analysis
26. What is Data Science? Also, list the differences between supervised and unsupervised learning.
27. Explain decision making statements in R.
28. Explain k-Nearest Neighbours (k-NN) with example.
29. Explain how naïve bayes used for spam filtering with example.
30. Explain Clustering of Social-Network Graphs.
31. Differentiate between univariate, bivariate and multivariate analysis with example.

**(6 x 4 = 24 Marks)**

### **PART D**

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

32. Explain decision tree algorithm. How I can calculate entropy and information gain for attributes (features) of a dataset?
33. What is sampling? what are the different sampling techniques?
34. What is data visualization? Explain basic principles of data visualization.
35. Explain how APIs and different tools used for web scraping.

**(2 x 15 = 30 Marks)**

\*\*\*\*\*