Reg. No.															

B.Tech. DEGREE EXAMINATION, MAY 2024

Fifth to Seventh Semester

18CSE396T – DATA SCIENCE

(For the candidates admitted during the academic year 2018-2019 to 2021-2022)

Note:

(i) **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.

(ii) **Part - B & Part - C** should be answered in answer booklet.

Time: 3 hours Max. Marks: 100

	Marks	BL	CO	PO
PART – A (20 × 1 = 20 Marks) Answer **ALL** Questions				
1. Which of the following is not a part of the data science process? (A) Communication building (B) Operations (C) Model planning (D) Discovery	1	1	1	1
2. What is semi-structured data? (A) Huge in number and has inaccurate value (B) Very less in number (C) Contains accurate values (D) Contains the data of both structured and semi structured data	1	1	1	1
3. Which of the following are the components of data science? (A) Statistics (B) Data expertise (C) Data engineering (D) Visualization	1	1	1	2
4. What is the aim of the NoSQL? (A) NoSQL is not suitable for storing structured data (B) NoSQL databases allow storing non-structured data (C) NoSQL is a new data format to store large data sets (D) NoSQL provides an alternative to SQL databases to store textual data	1	1	1	2
5. Data analytics uses _____ to get insights from data. (A) Statistical figures (B) Numerical methods (C) Statistical methods (D) Qualitative methods	1	1	2	1
6. The robotic arm will be able to paint every corner of the automotive parts while minimizing the quantity of paint wasted in the process. Which learning technique is sued in this problem? (A) Supervised learning (B) Unsupervised learning (C) Reinforcement learning (D) Semi-supervised learning	1	1	2	1
7. What kind of distance metric(s) are suitable for categorial variables to find the closest neighbors? (A) Euclidean distance (B) Manhattan distance (C) Minkowski distance (D) Hamming distance	1	1	2	1

8. In the normal distribution the data values are clustered around the 1 1 2 1
   (A) Mode (B) Standard deviation
   (C) Mean (D) Median

9. Which of the following is true for a vector in R? 1 1 3 1
   (A) It is a homogeneous 1-dimensional DS (B) It is a homogeneous 2-dimensional DS
   (C) It is a heterogeneous 1-dimensional DS (D) It is a heterogeneous 2-dimensional DS

10. In R programming, the function class ( ) belongs to which library? 1 1 3 1
   (A) Stats (B) Base
   (C) Utils (D) Class

11. Which of the following is not a meta character of regex in data analytics? 1 1 3 1
   (A) [ ] (B) ^
   (C) { } (D) #

12. _____ are generic data objects of R which are used to store the tabular data which are made up of three principal components, the data, rows and columns. 1 1 3 3
   (A) Array (B) Factors
   (C) Data frames (D) Lists

13. To distinguish between signal having very close values, we need an instrument with 1 1 5 1
   (A) High accuracy (B) High resolution
   (C) High sensitivity (D) High linearity

14. The expected value of voltage across a resistor is 80 V. However, the measurement gives a value of 79 V then % error and relative accuracy are 1 1 5 3
   (A) 1.25%, 0.9875 (B) 12.5%, 0.09875
   (C) 0.125%, 0.009875 (D) 1.025%, 0.0009875

15. Which one of the following has the highest accuracy? 1 1 5 3
   (A) Standard resistance (B) Standard inductance
   (C) Standard capacitance (D) Standard mutual inductance

16. _____ is harmonic mean of precision and recall score. 1 1 5 1
   (A) Accuracy (B) F1-score
   (C) Precision (D) Recall

17. How many types of BI users are there? 1 1 6 2
   (A) 2 (B) 3
   (C) 4 (D) 5

18. This is an approach to selling goods and services in which a prospect explicitly agrees in advance to receive marketing information. 1 1 6 2
   (A) Customer managed relationship (B) Data mining
   (C) Permission marketing (D) One-to-one marketing

19.	_____ produces bivariate scatter plots or time-series plots (A) XY plot (B) DCT plot (C) Bar chart (D) BW plot	1	1	6	2
20.	For bar chart, _____ non-trivial methods exists for tables and arrays, documented at bar chart table (A) Scatter plot (B) Dot plot (C) XY plot (D) Scatter plot and XY plot	1	1	6	2

PART – B (5 × 4 = 20 Marks)
Answer **ANY FIVE** Questions

		Marks	BL	CO	PO
21.	Disseminate the different stages in data science project.	4	2	1	2
22.	Write the difference between structured and unstructured data.	4	2	1	2
23.	Brief about data discovery and operationalized phase in data analytics life cycle.	4	2	2	3
24.	Write short notes on data exploration and model selection with suitable examples.	4	2	2	3
25.	Write the R-program to create a list, access the list, modify the list and concatenation of the list.	4	2	3	2
26.	Evaluate the clustering model with suitable examples.	4	2	5	5
27.	Discuss about deploying models.	4	2	5	3

PART – C (5 × 12 = 60 Marks)
Answer **ALL** Questions

		Marks	BL	CO	PO
28. a.	Imagine you are leading a data science project for a retail company aiming to optimize their inventory management system. The company has provided you with historical sales data, inventory records and customer feedback. Your goal is to develop a predictive model that can forecast future demand for direct products accurately, thereby minimizing stock outs and overstock situations. Can you outline the key stages you would undertake in this data science project, from data acquisition to model deployment, highlighting the importance of each stage and the specific task involved?	12	3	1	3
	(OR)				
b.	You are a data scientist tasked by a retail company to create a model predicting future product demand, optimizing their inventory management. You have historical sales data, inventory records, and customer feedback at your disposal. Outline the key stages of this data science project, explain the importance of each stage and specific tasks involved.	12	3	1	2
29. a.	Explain about business domain, resources, problem framing, key stakeholders, analytics sponsors, initial hypothesis.	12	2	2	3
	(OR)				

b.	List and explain the common tools for model building with an example.	12	2	2	3
30. a.	Illustrate arrays and its types write a program for two-dimensional array using R.	12	3	3	2
	(OR)				
b.	Analyze the following techniques with suitable example (i) Null hypothesis testing (ii) Alternative hypothesis	12	3	3	3
31. a.	Illustrate with example for the mapping problems to machine learning.	12	3	5	3
	(OR)				
b.	Evaluate a linear regression model and predicting logistic regression.	12	3	5	3
32. a.	Write in detail about deploying and exporting R HTTP services.	12	2	5	2
	(OR)				
b.	Explain data visualization in detail.	12	2	6	5

* * * * *