Reg. No.															

B.Tech. DEGREE EXAMINATION, MAY 2024

Fourth Semester

18ECE271T – INTRODUCTION TO DATA SCIENCE

(For the candidates admitted from 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 the most important language for data science? | 1 | 1 | 1 | 1,12
   (A) Java (B) Ruby
   (C) R (D) C

2. A column is a _____ representation of data. | 1 | 2 | 1 | 1,12
   (A) Diagonal (B) Vertical
   (C) Top (D) Horizontal

3. Python is a _____ language. | 1 | 3 | 1 | 1,12
   (A) Scripting (B) SQL
   (C) Data (D) Storage

4. Which one of the following skill a data scientist should have? | 1 | 3 | 1 | 1,12
   (A) Statistics knowledge (B) C
   (C) C++ (D) Java

5. TSV mean in data collection | 1 | 2 | 2 | 1,5
   (A) Tell Seperated Values (B) Ten Seperated Values
   (C) Tab-Seperated Values (D) Tab-Seperated Data

6. XML was designed to be _____readable. | 1 | 2 | 2 | 1,5
   (A) Human (B) Human and machine
   (C) Machine (D) Non

7. Removing noise from the data is called as | 1 | 3 | 2 | 1,5
   (A) Aggregation (B) Smoothing
   (C) Normalization (D) Generalization

8. Lack of symmetry is called | 1 | 1 | 2 | 1,5
   (A) Skow (B) Sew
   (C) Skew (D) Snow

9. When we don't known the labels from the training examples, we are using_____ method to learn 1 1 3 1,5

(A) Supervised learning (B) Unsupervised learning
(C) Reinforced learning (D) B and C

10. In statistics, a target is called as a _____ variable. 1 2 3 1,5

(A) Independent (B) Dependent
(C) Feature (D) Label

11. A transformation in statistics is called _____ in machine learning. 1 2 3 1,5

(A) Label creation (B) Feature
(C) Feature creation (D) Label

12. Machine learning algorithms are organized into a _____ 1 2 3 1,5

(A) Feature (B) Boxanomy
(C) Taxonomy (D) Texonomy

13. An _____question is constructed so the respondent cannot answer 'Yes' or 'No' but must give more information. 1 2 4,5 1,5

(A) Rank order type question (B) Close ended question
(C) Open ended question (D) Ended question

14. _____widely used in scaling responses in survey. 1 2 4,5 1,5

(A) A likert scale (1 to 5) (B) A likert scale (1 to 3)
(C) A likert scale (1 to 10) (D) A likert scale (1 to 20)

15. A good method for collecting data is _____ 1 2 4,5, 1,5

(A) Discussion (B) Surveys
(C) Data collection (D) Decision

16. Interview and focus group can deliver _____targeted information. 1 2 4,5 1,5

(A) Bad (B) Good
(C) Rich (D) Unknown

17. BIC for comparing model performance is _____ 1 3 6 1,5

(A) Baysian insight criterion (BIC) (B) Basic information criterion (BIC)
(C) Bayesian information criterion (BIC) (D) Bayesian information case (BIC)

18. The test sample is smaller data set compared to the _____ 1 2 6 1,5

(A) Training set (B) Model set
(C) Sample set (D) Model sample set

19. The formula to find recall is 1 3 6 1,5

(A) $R = \frac{PP}{TP + FN}$ (B) $R = \frac{TP}{PP + AN}$

(C) $R = \frac{FP}{TP + AN}$ (D) $R = \frac{TP}{TP + FN}$

20.	Cross-validation also called as _____ (A) Rotation estimation (B) Testing (C) Circular estimation (D) Validation	1	2	6	1,5

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

		Marks	BL	CO	PO
21.	Explain in short about computational skills.	4	2	1	1,12
22.	Describe about the tools used for data science.	4	2	1	1,12
23.	What are the challenges with unstructured data in data science?	4	3	2	1,12
24.	How to handle the missing data?	4	3	2	1,12
25.	Write short notes about gradient descent.	4	2	3	1,5
26.	Summarize about surveys.	4	2	4,5	1,5
27.	What is mean testing?	4	2	6	1,5

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

		Marks	BL	CO	PO
28. a.	Demonstrate how data science used in engineering and non-engineering domain. Also describe the impact of data science in engineering and non engineering domain.	12	2	1	1,12
	(OR)				
b.	Illustrate the skills required for a data science engineer. Explain in detail.	12	2	1	1,12
29. a.	Describe about the different process in data pre-processing in detail.	12	3	2	1,5
	(OR)				
b.	Illustrate the following with example (i) Predictive and perspective analytics (ii) Mechanistic analysis	6 6	3	2	1,5
30. a.	Describe about the following in detail (i) Regression: logistic (ii) Supervised learning	8 4	3	3	1,5
	(OR)				
b.	Illustrate about the following in detail (i) Softmax regression (ii) Machine learning	4 8	3	3	1,5

31. a.	Explain about the following terms					
	(i) Surveys	4				
	(ii) Survey question types	4	2	4,5	1,5	
	(iii) Survey audience	4				
	(OR)					
b.	Write about the interview (or) focus group data collection procedure and analyzing interview data in detail.	12	2	4,5	1,5	
32. a.	Describe about the following in detail					
	(i) Cross validation	6				
	(ii) Testing and A/B testing	6	2	6	1,5	
	(OR)					
b.	Write about the comparing models with necessary explanation.	12	2	6	1,5	

* * * * *