Reg. No.															

B.Tech. DEGREE EXAMINATION, NOVEMBER 2023

Fourth Semester

18ECE271T– INTRODUCTION TO DATA SCIENCE

(For the candidates admitted from the academic year 2020-2021 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. Data literacy computational thinking problem formulation is also known as | 1 | 2 | 1 | 1,12

(A) Abstraction (B) Automation
(C) Analyse (D) Completion

2. Python is a _____ Language. | 1 | 3 | 1 | 1,12

(A) Scripting (B) SQL
(C) Data (D) Storage

3. Vertical is a representation of what? | 1 | 2 | 1 | 1,12

(A) Rows (B) Column
(C) Both (A) and (B) (D) Degree

4. Data that endowed with meaning an purpose is | 1 | 1 | 1 | 1,12

(A) Information (B) Science
(C) Data (D) Data Collection

5. _____ data refers to highly organize information that can be seamlessly included in data base. | 1 | 3 | 2 | 1,5

(A) Un structured data (B) Structured data
(C) Both (A) and (B) (D) Under data

6. Data complication and organization is energy and time consuming task | 1 | 2 | 2 | 1,5

(A) Structured data (B) Data
(C) Un Structured data (D) Cleade Data

7. The abbreviation of CSV is | 1 | 2 | 2 | 1,5

(A) Call sell values (B) Column separate values
(C) Cell separated values (D) Comma separated values

8. During data Pre-Processing if data contains errors outliers then data is called as | 1 | 2 | 3 | 1,5

(A) Incomplete (B) Noisy
(C) Inconsistent (D) Good

9. A transformation in statistics is called _____ in machine learning. 1 2 3 1,5
   (A) Feature (B) Feature creation
   (C) Label (D) Future

10. When are the labels on the training examples known 1 1 3 1,5
   (A) Un supervised Learning (B) Supervised Learning
   (C) Both (B) and (A) (D) Reinforcement Learning

11. In machine learning a target is represented as _____ 1 2 3 1,5
   (A) Variable (B) Label
   (C) Size (D) Column

12. A variable in statistics is called as _____ in Machine Learning 1 3 3 1,5
   (A) Feature (B) Label
   (C) Creation (D) Values

13. _____ type questions ask respondents to choose one thing over another in preference. 1 3 4,5 1,5
   (A) MCQ Type (B) Rating Type
   (C) Order Type (D) Rank Order Type

14. _____ is widely used in scaling responses in surveys 1 2 4,5 1,5
   (A) A Likert Scale (1 to 3) (B) A Likert Scale (1 to 5)
   (C) A Likert Scale (1 to 10) (D) A Likert Scale (1 to 20)

15. An _____ question is constructed so that the respondent cannot answer 'yes' or 'no' But must give more information. 1 2 4,5 1,5
   (A) Rank order type Question (B) Ended Question
   (C) Close Ended Question (D) Open Ended Question

16. A survey can be conducted _____ , on paper, or _____ 1 1 4,5 1,5
   (A) In Person, Online (B) Mobile Contact, Online
   (C) Direct Question, Indirect Question (D) Open Ended, Close Ended

17. One of the metric to evaluate the goodness of the model. 1 2 6 1,5
   (A) Value (B) Feature
   (C) Precision (D) Quantity

18. The formula to find recall is 1 3 6 1,5
   (A) $\text{Re}call = \frac{PP}{TP \& FN}$ (B) $\text{Re}call = \frac{PP}{TP \& AN}$
   (C) $\text{Re}call = \frac{FP}{TP \& FN}$ (D) $\text{Re}call = \frac{TP}{TP \& AN}$

19. AIC in comparing model is 1 2 6 1,5
   (A) AIC (Akaike Information Criterion) (B) AIC (C Almond Information Criterion)
   (C) AIC (ALL Information Criterion) (D) AIC (Air Information Criterion)

20.	The error it makes in predictions are accumulated. This gives what is often called? (A) Median Absolute Test et error. (B) Non – Absolute Test Set Error. (C) Mean – Absolute Test set error. (D) Test Set Error	1	2	6	1,5

PART – B (5 × 4 = 20 Marks)

Answer **ANY FIVE** Questions

		Marks	BL	CO	PO
21.	Explain where do we see data science?	4	2	1	1,12
22.	Summarize about the relationship between data science and information science.	4	2	1	1,12
23.	Write about Open Data in Data Collections.	4	3	2	1,12
24.	Summarize the data storage and presentation.	4	2	2	1,12
25.	What are the types of regression and explain about Linear Regression in short.	4	3	3	1,5
26.	Describe about Pros and Cons of Surveys.	4	2	4,5	1,5
27.	What is Mean Quantitative Methods?	4	3	6	1,5

PART – C (5 × 12 = 60 Marks)

Answer **ALL** Questions

		Marks	BL	CO	PO
28. a.	Describe the following in detail.		2	1	1,12
	i. Computational Thinking	6			
	ii. Tools need for Data Science	6			
	(OR)				
b.	Illustrate the following in detail.		2	1	1,12
	i. Skills required for Data Science	6			
	ii. Issues of Ethics, Bias and Privacy in Data Science	6			
29. a.	Explain about the different data types in detail.	12	3	2	1,5
	(OR)				
b.	Describe about the data Collection in detail.	12	3	2	1,5
30. a.	Write about the different Types of Regression (Logistic and Softmax) with Example.	12	3	3	1,5
	(OR)				
b.	Describe about the Gradient Descent approach in detail.	12	3	3	1,5
31. a.	Describe the following in detail.		2	4,5	1,5
	i. Survey Question Types	6			
	ii. Analyzing Survey Data	6			
	(OR)				

b.	Explain about the interview (or) focus Group Procedure and Pros and Cons of interviews and Focus Group.	12	2	4,5	1,5
32. a.	Write about the comparing model in detail.	12	3	6	1,5
	(OR)				
b.	Explain about the Training, Testing and A/B Testing in detail with examples.	12	3	6	1,5

* * * * *