



GIET UNIVERSITY, GUNUPUR – 765022

B. Tech (Third Semester - Regular) Examinations, December – 2022

21BCSPE23011 / 21BCDPE23011 – Introduction to Data Science (CSE, CSE(DS))

Time: 3 hrs

Maximum: 70 Marks

Answer ALL questions

(The figures in the right-hand margin indicate marks)

PART – A**(2 x 5 = 10 Marks)**Q.1. Answer *ALL* questions

- | | CO # | Blooms Level |
|--|------|--------------|
| a. Explain Data science and differentiate between structured and unstructured data | 1 | 2 |
| b. Explore roles of Data Scientist. | 1 | 1 |
| c. Exemplify the importance of data cleaning in Pre-processing. | 2 | 2 |
| d. Explain term in-Sample evaluation of model and represent Mean square error. | 3 | 1 |
| e. Illustration the term Over fitting. | 4 | 1 |

PART –B**(15 x 4 = 60 Marks)**Answer ALL questions

- | | Marks | CO # | Blooms Level |
|---|-------|------|--------------|
| 2. a. Describe The NOIR scale of data classification. | 8 | 1 | 2 |
| b. Briefly describe the evolution of data science in different era. | 7 | 1 | 2 |
| (OR) | | | |
| c. Illustrate all the stages of Data science project Lifecycle with proper diagram. | 8 | 1 | 2 |
| d. Discusses various issue in data security. | 7 | 1 | 2 |
| 3.a. Calculate the standard deviation for the following data | 8 | 2 | 3 |

No of rejects per operator	21-25	26-30	31-35	36-40	41-45	46-50	51-55
No of Operators	5	15	28	42	15	12	03

- | | | | |
|---|---|---|---|
| b. With neat diagram describe the skewness in data distribution. | 7 | 2 | 3 |
| (OR) | | | |
| c. Based on the frequency distribution given below, evaluate coefficient of variance. | 8 | 2 | 3 |

Annual tax paid (Rs Thousand)	5-10	10-15	15-20	20-25	25-30	30-35	35-40
No of Operators	18	30	46	28	20	12	6

- | | | | |
|---|---|---|---|
| d. Define correlation. Classify corelation with examples. | 7 | 2 | 3 |
| 4.a. Define the term simple linear regression. Evaluate the regression from the given data and evaluate the standard error. | 8 | 3 | 3 |

X	1	3	10	16	26	36
Y	42	50	75	100	150	200

b. Describe different methods to evaluate the Regression model. 7 3 2

(OR)

c. Define the term multiple linear regression. Evaluate the regression line from the given data 8 3 3

X1	2	3	5	7	8
X2	1	5	3	6	7
Y	3	2	4	5	8

d. How residual plot helps for regression model validation. 7 3 2

5.a. Define Bias and variance. What is the need of Bias variance trade off. 10 4 1

b. Illustrate the importance of Ridge regression 5 4 2

(OR)

c. Describe different types of cross validation. 10 4 1

d. Define the term Grid search. 5 4 2

--- End of Paper ---