GIET UNIVERSITY, GUNUPUR – 765022

B. Tech (Third Semester) Examinations, December – 2023

21BCSPE23011 /22BCSPE23011– Introduction to Data Science

(CSE,CSE(DS))

Time: 3 hrs Maximum: 70 Marks

(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. Define Data science. What is the difference between structured and unstructured data.	CO1	K1
b. Differentiate between Ordinal and Ratio type of data.	CO1	K4
c. If the mean of a dataset is 50 and the standard deviation is 10, interpret this in the context of the data.	CO2	K3
d. Define the procedure for finding the Standard Error for a dataset.	CO3	K1
e. Explain the concept of k-Fold Cross Validation.	CO4	K2

PART – B **(15 x 4 = 60 Marks)**

Answer ***ALL*** questions	Marks	CO #	Blooms Level
2. a. Illustrate all the stages of Data science project Lifecycle with proper diagram.	12	CO1	2
b. Discuss the role of data science in the field of Education.	3	CO1	2
(OR)			
c. Explain the fundamental principles of data security, and why are they important in the digital age? What are the major threats to data security.	10	CO1	2
d. .Explain the different ways of collection of data.	5	CO1	2
3.a. With neat diagram describe the skewness in data distribution.	7	CO2	1
b. Based on the frequency distribution given below, evaluate coefficient of variance.	8	CO2	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

(OR)

c. Describe normalization. Describe the different methods of normalization.	7	CO2	1
d. Evaluate the Kerl Pearson measure of skewness in basis of Mean, mode and standard deviation from the following data	8	CO2	3

Class interval	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22
Frequency	40	35	46	98	125	87	45	24

4.a. Explain Residual plot and scatter plot with proper representation. 5 CO3 1

b. Define the term simple linear regression. Evaluate the regression from the given data and evaluate the standard error. 10 CO3 3

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

(OR)

c. Differentiate between Linear Regression and Polynomial Regression. 5 CO3 4

d. Describe the importance of Polynomial regression. Find Polynomial regression of degree two from the given data. 10 CO3 3

X	1	3	4	7	9
Y	1	6	1	8	20

5.a. Define Bias and variance. What is the need of Bias variance trade off. 10 CO4 2

b. Discuss about the train and test sample set in a dataset. What are its application in a model? 5 CO4 2

(OR)

c. How does Ridge Regression contribute to stable and reliable predictions in the presence of noise in the data? Explain with example. 10 CO4 2

d. How model error is different from generalized error. 5 CO4 4

--- End of Paper ---