BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: BCA **SEMESTER : IV**
BRANCH: BCA **SESSION : SP/2025**

SUBJECT: CN227 INTRODUCTION TO DATA SCIENCE

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
***4*. Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	Define Data Science and describe different steps of data science process.	[5]	1	2
Q.1(b)	Explain the Big Data Eco System and its components.	[5]	1	2
Q.2(a)	Describe the different types of data with example used in Data Science.	[5]	1	2
Q.2(b)	Explain the different steps of data preprocessing.	[5]	1	2
Q.3(a)	In a survey of 950 families in a village, the following distribution of number of children was obtained:	[5]	3	3

No. of children	0 - 2	2 - 4	4 - 6	6 - 8	8 – 10	10 - 12
No. of families	272	328	205	120	15	10

Find Standard Deviation.

Q.3(b)	What is Heat Map and how to interpret it? Create a box plot of the following sample data: 120, 140, 130, 170, 130, 160, 150, 190, 140, 250, 160, 130	[5]	3	3
Q.4(a)	Calculate the two regression equations of *X* on *Y* and *Y* on *X* from the data given below, taking deviations from an actual means of *X* and *Y*.	[5]	4	3

Price(Rs.)	10	12	13	12	16	15
Amount Demanded	40	38	43	45	37	43

Estimate the likely demand when the price is Rs.20.

Q.4(b)	Explain the evaluation of model through Confusion matrix for Binary Class with example.	[5]	4	2
Q.5(a)	Explain test statistic from sample data in hypothesis testing with population parameters.	[5]	5	2
Q.5(b)	Define cross-validation, overfitting and underfitting with examples.	[5]	5	2

::::::01/05/2025::::::M