Enrollment Number: _____ Roll Number: _____

JAMIA HAMDARD
Department of Computer Science & Engineering
School of Engineering Sciences & Technology
New Delhi – 110062

Examination: Odd Semester Examination 2022

Sub: Data Science Sub. Code: BTCSE – DED51 / BTCSE – DED51 (Lateral Entry)

Max. Time: 3 hours Max. Marks: 75

Note: Follow proper numbering for answering the questions.

SECTION – A

Attempt all questions. *7 x 5 = 35 Marks*

SL. NO	Questions	Marks	CO	BL
1.	What do you mean by Data Science? Discuss some applications of Data Science.	5	CO1	L1
2.	Explain the 4 V's of Big Data.	5	CO2	L1
3.	Differentiate the following: i. Clustering vs Classification ii. Supervised learning vs Unsupervised learning.	5	CO3	L4
4.	Justify the statement "Deep learning is highly beneficial to Data Science." Give a suitable real-life example in support of your answer.	5	CO4	L5
5.	Explain how Data visualization is important for Data Science.	5	CO3	L2
6.	Discuss the various types of data used in data science.	5	CO1	L2
7.	Illustrate a scenario where: i. Supervised learning gives better results when compared to unsupervised learning. ii. Unsupervised learning gives better results when compared to supervised learning.	5	CO3	L6

SECTION – B

Attempt any four out of six questions. *4 x 10 = 40 Marks*

SL. NO	Questions	Marks	CO	BL
1.	Use a suitable real-life example to explain all steps of data science process.	10	CO1	L2 L6
2.	The blood pressures, p mmHg, and the ages, t years, of 7 hospital patients are shown in the table below:	10	CO2	L3

Patient	A	B	C	D	E	F	G
t	42	74	48	35	56	26	60
p	98	130	120	88	182	80	135

i. Find the equation of the regression line of p on t.
ii. Use the model to find the blood pressure of a patient with age 40.
iii. Also, check the fitness of the model.

3.	What is machine learning and explain the modeling process. Also discuss how training, testing and validation of model cruicial for machine learning	10	CO3	L1 L4
4.	Discuss the limitations of machine learning and thus explain the necessity to move towards deep learning. Also write some applications of deep learning.	10	CO4	L2 L4
5.	Write a short note on i. Data visualization options. ii. Data visualization tools	10	CO5	L2
6.	Use K-means clustering, and distance given by the formula D to cluster the following 10 points into 3 clusters. Show the calculation performed at each iteration. Assume that the initial clusters are A, E and H. $D(X,Y) = \|\|x2 - x1\| + \|y2 - y1\|\|$ Where X is (x1, y1) and Y is (x2, y2).	10	CO3	L3

Point	X	Y
A	2	10
B	2	5
C	8	4
D	5	8
E	4	9
F	6	4
G	1	2
H	7	5
I	3	7
J	9	3

BL – Bloom's Taxonomy Levels (1 – Remembering, 2 – Understanding, 3 – Applying, 4 – Analyzing, 5 – Evaluating, 6 – Creating)

****