

**BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(MID SEMESTER EXAMINATION)**

**CLASS: B.TECH.
BRANCH: BT/CIVIL/ECE/PROD/MECH**

**SEMESTER: VII
SESSION: MO/2022**

SUBJECT: IT340 MACHINE LEARNING

TIME: 2 HOURS

FULL MARKS: 25

INSTRUCTIONS:

1. The total marks of the questions are 25.
 2. Candidates attempt for all 25 marks.
 3. Before attempting the question paper, be sure that you have got the correct question paper.
 4. The missing data, if any, may be assumed suitably.
 5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.
-

			CO	BL
Q1	(a) Define univariate and bivariate analysis.	[2]	CO2	1
Q1	(b) Why is data pre-processing necessary? What are the different measures to perform cleaning of data?	[3]	CO1, CO3	2
Q2	(a) Contrast and compare supervised and unsupervised machine learning approach.	[2]	CO2, CO3	4
Q2	(b) What do you mean by categorical attribute? Illustrate the types of categorical attribute using example.	[3]	CO2	1
Q3	(a) The following observations are arranged in ascending order. The median of the data is 25. Evaluate the value of x. 17, x, 24, x+7, 35, 36, 46	[2]	CO1	3
Q3	(b) Why is data transformation required? Use max-min normalization (in the range of 0 to 1) to normalize the values 200, 300, 400, 600, 1000.	[3]	CO1, CO2	2, 3
Q4	(a) What strategies can be implemented to minimize the overfitting problems in decision tree?	[2]	CO2	2
Q4	(b) Why is cost function required in regression? Give mathematical expression of cost function.	[3]	CO4	4
Q5	(a) Why tree pruning is necessary?	[2]	CO3	2
Q5	(b) Use the following data set to find the information gain [Info Gain(Age)] for the attribute age. Risk represents the class label [H means 'High' and L means 'Low'], Age : Y is Young, M is Middle Age, S is Senior.	[3]	CO3	3

ID	Age	Car	Risk
1	M	Sports	L
2	Y	Vintage	H
3	M	Sports	L
4	S	SUV	H
5	Y	Sports	H
6	M	SUV	H

:::::: 01/10/2022 :::::M