MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
A Constituent Institution of Manipal University

II SEM. M.C.A IN-SEMESTER EXAMINATIONS June 2021
SUBJECT: DATA ANALYTICS [MCA 4251]

Date of Exam: **07/06/2021**   Time of Exam: **10:30 am– 12:30 noon**   Max. Marks: **30**

Instructions to Candidates:

❖ Answer ALL the questions & missing data may be suitable assumed
❖ Use of calculators is permitted

1	Distinguish between dichotomous and nominal data variables using appropriate examples.	2
2	What is the role and responsibilities of a subject matter expert in data analysis projects?	2
3	How is a Contingency table different from a Summary table? Give examples.	2
4	Explain any two strategies to separate test data set from training data set for the purpose of calculating accuracy of classification or prediction.	2
5	What are outliers? Mention any two strategies to deal with outliers in datasets.	3
6	A binary classifier predicts 1 (positive case) and 0 (negative case). A separate test set of 20 examples is used to test the model and the results are available in the table below. Calculate the model's accuracy measures: ➢ Concordance ➢ Sensitivity ➢ Specificity	3

Observation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actual	0	1	1	0	0	1	0	0	1	1	1	0	0	1	0	1	0	1	0	0
Predicted	0	1	1	0	0	0	0	0	1	1	1	1	0	1	0	1	0	1	1	0

7	Compare and Contrast Descriptive and Predictive Analytics with an example for each.	3

8	The following table shows the relationship between the amount of fertilizer used and the height of a plant.
	   i. Calculate a simple linear regression equation using Fertilizer as the descriptor and Height as the response.
	   ii. Predict the height when fertilizer is 8.5.
	   iii. Visualize using a scatter plot.
		4

Fertilizer	10	5	11	18	14	7	14	13	7	8	10	11	16	20	17
Height	0.8	0.4	0.8	1.4	1.1	0.6	1.3	1.1	0.6	0.7	0.7	0.9	1.3	1.7	1.3

9	A leading fashion store chooses to predict the willingness of a customer to buy a shirt of a particular price category based on the customers' data. The company strongly believes that the willingness of a customer to buy depends on 3 factors – gender, the type of car used by the customer and the shirt price category. Use the Naïve Bayesian classifier method to determine if a customer would buy a shirt if gender="Male", car type="sports", shirt price category ="expensive".	4

Customer ID	Gender	Car Type	Shirt price category	Will Buy?
1	Male	Sports	Cheap	No
2	Male	Sports	Expensive	Yes

	3	Male	Family	Cheap	Yes		
	4	Male	Family	Expensive	No		
	5	Male	Sports	Cheap	Yes		
	6	Male	Sports	Expensive	Yes		
	7	Male	Family	Cheap	Yes		
	8	Male	Family	Expensive	No		
	9	Female	Sports	Cheap	No		
	10	Female	Family	Cheap	No		
	11	Female	Sports	Expensive	No		
	12	Female	Family	Expensive	No		
10	(a) Perform Agglomeration using single link and Euclidean distance Measure on the following data set. (b) Draw the dendrograms for the Agglomerative clustering. (c) Draw the step-wise scatter plots to visualize the clustering.						**5**

	X	Y
P1	10	15
P2	12	18
P3	15	22
P4	19	27
P5	24	33
P6	30	40