

**COLLEGE OF ENGINEERING TRIVANDRUM DEPARTMENT OF
COMPUTER APPLICATIONS**

Fifth Semester M.C.A Degree

First Series Examination January-2022

20MCA201: DATA SCIENCE & MACHINE LEARNING

Time: 1 hr

Max. Marks: 20

PART-A

(Answer All Questions. Each question carries 2.5 marks)

1. Do you think that Correlation and Causation are same? Justify your answer.
2. Three bulbs are chosen at random from 15 bulbs, of which 5 are defective.

Find the probability that

- (i) none is defective
- (ii) exactly one is defective
- (iii) atleast one is defective

3. In a city, 51% of the adults are males. One adult is randomly selected for a survey involving credit card usage.
 - (a) Find the prior probability that the selected person is a female.
 - (b) It is later learned that the selected survey subject was smoking a cigar. Also, 9.5% of males smoke cigars, whereas 1.7% of females smoke cigars. Use this additional information to find the probability that the selected person is a female.
4. With a neat plot explain an univariate method to visualize data.

(Total : 4 x 2.5= 10 marks)

PART-B

(Each question carries 5 marks)

5. Based on a survey conducted in an institution, students are classified based on the two attributes of academic excellence and other activities. Given the following data, identify the classification of a student with $X = 5$ and $Y = 7$ using k-NN algorithm (choose k as 3).

X (Academic Excellence)	Y (Other Activities)	Z (Classification)
8	6	Outstanding
5	6	Good
7	3	Good
6	9	Outstanding

OR

6. Consider the set of training samples given below .Use Bayesian model to test the classifier for a new instance $X = \{\text{Outlook} = \text{Sunny}; \text{Temperature} = \text{Cool} ; \text{Humidity} = \text{High} ; \text{Wind} = \text{Strong}\}$.

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

7. Discuss the following multivariate visual exploration techniques with a neat plot

- (a) Scatter Plot

(c) Density Chart

(b) Bubble Chart

(d) Distribution Chart

OR

8. Explain various processes for preparing a data set.

(Total: 2 x 5= 10 marks)

OBE Mapping Table		
Question Number	Cognitive Knowledge Level in Bloom's Taxonomy *	Course Outcome
2,3,5,6	B3	CO2 Explain the basics of machine learning and use lazy learning and probabilistic learning algorithms to solve data science problems.
1,4	B2	CO2
* Remember – B1, Understand – B2, Apply – B3, Analyse – B4, Evaluate – B5, Create - B6		

Hardness Level	Question Numbers	Cumulative Marks	% on Total Marks (count the marks for choice questions)
Average	1, 4,5	10	50
Medium	6,2	7.5	40
Hard	3	2.5	10
\$ - Average (50% - 60%), Medium (30% to 40%) and Hard (10%)			