Reg No.:_____ **0520MCA201122203** Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Third Semester MCA (Two Year) Degree (R,S) Examination December 2024

Course Code: 20MCA201

Course Name: DATA SCIENCE AND MACHINE LEARNING

Max. Marks: 60 Duration: 3 Hours

PART A

Answer all questions, each carries 3 marks. Marks

1	Discuss the significance of Data Science.	(3)
2	What is Quartile plot? Explain with an example.	(3)
3	Briefly discuss different types of learning algorithms.	(3)
4	Explain the role of Laplace estimator in Bayesian classification.	(3)
5	Differentiate between entropy and information gain.	(3)
6	What is Pearson's coefficient of correlation? Discuss its significance in data analysis.	(3)
7	With a neat diagram, explain the concept of artificial neurons (Perceptrons) in neural networks.	(3)
8	What is convex hull in SVM? Explain.	(3)
9	Differentiate between sensitivity and specificity.	(3)
10	Explain bootstrap sampling.	(3)

PART B

Answer any one question from each module. Each question carries 6 marks.

Module I

11	Discuss the significance of data visualization. Explain the following visualization techniques. (i) Histogram (ii) Scatter Plot (iii) Density Chart	(6)

OR

12 With a suitable diagram, explain the various steps involved in a data science process. (6)

Module II

13 With the given data, Use k-NN algorithm to determine the Taget attribute for a new instance with X = 5 and Y =3. (Choose k as 3) (6)

X	Y	TARGET
2	3	Class1
3	3	Class1
2	8	Class2
4	1	Class1
5	8	Class2
6	7	Class2

OR

14 Consider the training data of 10 samples in the given table where 'Play' is a class attribute. Use Bayesian classifier to predict whether there will be a play if it is a rainy day with mild temperature, Normal humidity and Strong wind. (6)

Day	**Outlook**	**Temperature**	**Humidity**	**Wind**	**PLAY**
1	Sunny	Hot	High	Weak	No
2	Sunny	Cool	Normal	Weak	Yes
3	Overcast	Hot	High	Weak	Yes
4	Sunny	Mild	High	Weak	No
5	Sunny	Mild	Normal	Strong	Yes
6	Rain	Mild	High	Weak	Yes

7	Rain	Cool	Normal	Weak	Yes
8	Rain	Cool	Normal	Strong	No
9	Sunny	Hot	High	Strong	No
10	Overcast	Mild	High	Strong	Yes

Module III

15 List the benefits of pruning in decision trees? Explain various approaches to tree pruning? (6)

OR

16 Distinguish between classification and regression with suitable examples (6)

Module IV

17 Explain in detail Back Propagation Learning algorithm. (6)

OR

18 What is kernel trick in SVM? Discuss its significance for non-linearly separable data. (6)

Module V

19 What are ROC space and ROC curve in machine learning? In ROC space, which points correspond to perfect prediction, always positive prediction and always negative prediction? Why? (6)

OR

20 Discuss various ways of improving the performance of learning models (6)

****