

Register No.: ..... Name: .....

**SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)**

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

**THIRD SEMESTER MCA DEGREE EXAMINATION (R), DECEMBER 2023****(2021 SCHEME)****Course Code: 21CA301****Course Name: Data Science and Machine Learning****Max. Marks: 60****Duration: 3 Hours****PART A*****(Answer all questions. Each question carries 3 marks)***

1. What is clustering?
2. Explain the concept data preparation.
3. What is the relevance of a model in the knowledge representation of a machine?
4. Explain the following terms:
  - i. Posterior Probability
  - ii. Prior Probability
  - iii. Marginal Likelihood.
5. Describe ordinary least squares estimation.
6. What is entropy?
7. Explain threshold activation function.
8. List any three applications of SVM (Support Vector Machine).
9. What is an ensemble learning?
10. Compare bagging and boosting.

**PART B*****(Answer one full question from each module, each question carries 6 marks)*****MODULE I**

11. Explain the terms.
  - i. Outliers
  - ii. Data Sampling (6)
  - iii. Measure of Central Tendency
  - iv. Measure of Spread

**OR**

12. How do decision trees function as a machine learning algorithm? What are the fundamental principles and techniques involved in constructing the decision tree models to make accurate predictions or classifications in different domains? (6)

## MODULE II

13. Consider the following datasets with attributes height, weight and class labels are underweight and normal. Suppose that a new observation is added with height 170 and weight 57. Find class label of the new observation.

Height (in cm)	Weight (in cm)	Class Label
167	51	Underweight
182	62	Normal
176	69	Normal
173	64	Normal
172	65	Normal
174	56	Underweight
169	58	Normal
173	57	Normal
170	55	Normal
170	57	?

(6)

OR

14. Write Naïve Bayes algorithm. Find out the play prediction "yes" or "no" when the weather is sunny.

NO:	Outlook	Play
1	Rainy	Yes
2	Sunny	Yes
3	Overcast	Yes
4	Overcast	Yes
5	Sunny	No
6	Rainy	Yes
7	Sunny	Yes
8	Overcast	Yes
9	Rainy	No
10	Sunny	No
11	Sunny	Yes
12	Rainy	No
13	Overcast	Yes
14	Overcast	Yes

(6)

## MODULE III

15. Describe the divide and conquer approach used in constructing decision trees for machine learning. Illustrate it with a specific example or scenario to demonstrate how this method helps in making optimal decisions at each node of the tree.

(6)

OR

16. a) Explain simple linear regression and multiple linear regression. How does it differ?

(4)

- b) When do you say that a decision tree has been pruned? (2)

**MODULE IV**

17. Briefly explain the concept of artificial neural networks (ANN). (6)

**OR**

18. What does Maximum Margin Hyperplane (MMH) mean? When the data are linearly separable, how do you determine MMH? (6)

**MODULE V**

19. Let there be ten balls (six white and four red balls) in a box and let it be required to pick up the red balls from them. Suppose we pick up seven balls as the red balls, of which only two are actually red balls. What are the values of precision, recall, accuracy, F-measure, specificity and error rate in picking a red ball? (6)

**OR**

20. Explain in detail the cross validation methods. (6)

\*\*\*\*\*