

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

**SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)**

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

**THIRD SEMESTER M.C.A DEGREE EXAMINATION (S), MAY 2022****(2020 SCHEME)****Course Code: 20MCAT201****Course Name: Data Science & Machine Learning****Max. Marks: 60****Duration: 3 Hours****PART A***(Answer all questions. Each question carries 3 marks)*

1. Compare traditional programming and machine learning.
2. Explain the desired format of datasets in data science.
3. Why is the k-NN classification considered as lazy learning algorithm?
4. Give the relevance of joint probability and conditional probability for Bayesian classifier.
5. How do you choose the best split based on the decision tree algorithm?
6. Explain the use of correlation in linear regression.
7. What is epoch in back-propagation algorithm?
8. How do you choose the maximum margin hyperplane?
9. Write the K-means clustering algorithm.
10. How does confusion matrix help in performance evaluation?

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

11. Explain with a neat diagram the classification of data science problems into different tasks. (6)

**OR**

12. Descriptive statistics plays a fundamental role in exploratory data analysis. Write about the measures of central tendency and the measures of spread with proper examples. (6)

**MODULE II**

13. Explain the k-NN classification algorithm with the help of a suitable example of your choice. (6)

**OR**

14. a) Write Naïve Bayes Classification algorithm. (2)  
b) Consider the car theft problem with attributes color, type, origin, and the target, stolen can be either Yes or No. Dataset is expressed below. (4)

Classify (red, SUV, domestic) using Naïve Bayes classifier.

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

### MODULE III

15. Describe C5.0 decision tree algorithm with suitable example. (6)

OR

16. Explain simple linear regression in detail. (6)

### MODULE IV

17. Discuss any five activation functions of neural networks with appropriate plots. (6)

OR

18. How does SVM handle non-linearly separable data using kernels? (6)

### MODULE V

19. What is hierarchical clustering? Write the agglomerative hierarchical clustering algorithm. (6)

OR

20. Consider the confusion matrix given below for a binary classifier predicting the presence of a disease. (6)

	Predicted No	Predicted Yes
Actual No	45	5
Actual Yes	5	95

The classifier made a total of 150 predictions. Out of those 150 cases, the classifier predicted "yes" 100 times, and "no" 50 times. In reality, 100 patients in the sample have the disease, and 50 patients do not. Calculate the following terms from the given confusion matrix. (6)

1. Precision
2. Recall
3. Accuracy.

\*\*\*\*\*