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

SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

THIRD SEMESTER M.C.A DEGREE EXAMINATION (Regular), DECEMBER 2022
(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. How do you find a five-number summary?
2. How does uniform distribution differ from normal distribution?
3. Discuss the learning process of a machine.
4. Give the relevance of supervised learning in machine learning.
5. Explain the divide-and-conquer technology / strategy.
6. What is the relevance of information gain?
7. Write a short note on unit step activation function.
8. What is a perceptron?
9. How does SVM handle non- linearly separable data?
10. How does boosting process improve model performance?

PART B

(Answer one full question from each module, each question carries 6 marks)

MODULE I

11. With a neat diagram, explain CRISP - data mining framework. (6)

OR

12. Explain multivariate visualization and univariate visualization. (6)

MODULE II

13. With an example, explain Naive Bayes classification algorithm. (6)

OR

14. Suppose the age and gender of a group of people and their corresponding class of sports are given. Find the class of sports of Rohith whose age and gender are respectively 5 and 0. (use KNN) (6)

Name	Gender	Age	Class Label
Aleena	1	32	Football
Abdul	0	16	Handball
Akshay	0	34	Basketball
Sara	1	55	Basketball
Kajol	1	40	Handball
Dushiyant	0	20	Handball
Pihu	1	15	Handball
Arman	0	55	Football

MODULE III

15. a) Obtain a linear regression for the data given in the table below assuming that y is the independent variable. (4)

Invest(X)	Outcome(Y)
5	5
4	4
3	5
2	4
1	2

b) Discuss the strengths and weaknesses of decision tree algorithm (C5.0 algorithm). (2)

OR

16. What do you mean by decision trees? With the help of an example, explain the divide and conquer method for the construction of decision trees. (6)

MODULE IV

17. What are activation functions? Discuss any three types of activation functions. (6)

OR

18. Explain the backpropagation algorithm for training the given data using artificial neural networks. State the strengths and weaknesses of such multilayer feedforward networks. (6)

MODULE V

19. How will you evaluate the performance of a model using confusion matrix? (6)

OR

20. A classification model is used to predict whether a person would default on a bank loan. To build this classification model, let's say, a historical data set of 10000 records got chosen for building the model. As part of building the model, all of the 10000 records got labeled where each record represented a person and got labeled as "Yes" or "No" based on whether they defaulted (Yes) or not defaulted (No). Explain and find the accuracy, precision, and recall. (6)

****