LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION – **STATISTICS**

THIRD SEMESTER – **NOVEMBER 2023**

PST3MC03 – DATA MINING AND MACHINE LEARNING

Date: 04-11-2023   Dept. No.   Max. : 100 Marks
Time: 01:00 PM - 04:00 PM

SECTION A – K1 (CO1)

	Answer ALL the questions	(5 x 1 = 5)
1	**Define the following**	
a)	Spatial databases.	
b)	Mining methodology issues.	
c)	Parameters in Random forest.	
d)	Assumption of Naïve Bayesian classifier.	
e)	Performance of Machine Learning algorithms in terms of their error rate.	

SECTION A – K2 (CO1)

	Answer ALL the questions	(5 x 1 = 5)
2	**Answer ALL the questions**	
a)	Decision trees are a type of ------------------ learning algorithm used for both classification and regression tasks.	
b)	Entropy impurity measure ----------------------------- .	
c)	Random Forest introduces --------------- sampling of features during the tree-building process.	
d)	The term ----------------- refers to the number of layers in a neural network between the input and output layers.	
e)	Boosting algorithms aim to reduce both ------------- and variance in the final ensemble model.	

SECTION B – K3 (CO2)

	Answer any THREE of the following	(3 x 10 = 30)
3	What type of data are used in Machine Learing Algorithm?	
4	Describe any five application of Data mining.	
5	Explain Accuracy, Error rate, Sensitivity, Specificity, F-score of Machine learning algorithm.	
6	Examine the steps involved in the Apriori algorithm for association rule mining.	
7	What are the advantages and disadvantages of the Decision tree algorithm?	

SECTION C – K4 (CO3)

	Answer any TWO of the following	(2 x 12.5 = 25)
8	How does K^{th} Nearest Neighbourhood algorithm work with weighted and unweighted approach?	
9	State the fundamental distinctions and resemblances between classification and regression trees algorithms.	
10	Using the Naïve Bayesian algorithm, what is the predicted outcome ("play") when humidity is high, the outlook is sunny, and it is windy (false), based on the provided dataset?	

1

outlook	humidity	windy	play
rain	moderate	TRUE	TRUE
rain	moderate	TRUE	FALSE
rain	moderate	TRUE	FALSE
sunny	moderate	TRUE	FALSE
rain	moderate	FALSE	TRUE
rain	moderate	FALSE	TRUE
sunny	moderate	FALSE	FALSE
rain	moderate	TRUE	TRUE
rain	moderate	TRUE	TRUE
rain	moderate	FALSE	TRUE
sunny	high	TRUE	FALSE
sunny	high	TRUE	FALSE
rain	high	TRUE	FALSE
rain	high	TRUE	TRUE
rain	high	TRUE	TRUE
sunny	high	TRUE	TRUE
rain	high	TRUE	FALSE
rain	high	TRUE	FALSE
sunny	high	FALSE	TRUE
sunny	high	FALSE	TRUE
sunny	high	FALSE	TRUE
rain	high	FALSE	TRUE

11	Write the algorithm of Adaptive Boosting for classification problem.

SECTION D – K5 (CO4)

Answer any ONE of the following	**(1 x 15 = 15)**

12	Explain the steps involved in constructing a classification tree using algorithmic procedures.
13	What factors contribute to the current popularity of data mining?

SECTION E – K6 (CO5)

Answer any ONE of the following	**(1 x 20 = 20)**

14	Elaborate on the steps involved in the Random Forest algorithm for solving regression tasks.
15	How does the Artificial Neural Network work for the classification problem using back propagation method?

&&&&&&&&&&