

B.TECH/CSE/EVEN/8th Sem/R18/ CS802B/2021-2022
YEAR: 2022

MACHINE LEARNING

PAPER CODE: CS802B

TIME ALLOTTED: 3 HOURS

FULL MARKS: 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable

GROUP – A

(Multiple Choice Type Questions)

1. Answer any **ten** from the following, choosing the correct alternative of each question: **10×1=10**

SL	Question	Marks	CO	Blooms Taxonomy Level
(i)	During training phase, dataset with target class labels is required in (a) Supervised Learning (b) Unsupervised Learning (c) Both Supervised and Unsupervised Learning (d) None of the above	1	CO2	BLT1
(ii)	A similarity distance criterion is to be decided in (a) Clustering (b) Association Rule Mining (c) Classification (d) None of the above	1	CO3	BLT2
(iii)	Regression Model is used to predict (a) Numerical value of the target variable (b) Categorical class of target variable (c) Similarity between the target variable and the predictor variable (d) None of the above	1	CO2	BLT1
(iv)	Market Basket analysis is based on the concept of (a) Clustering (b) Association Rule Mining (c) Classification (d) None of the above	1	CO3	BLT1
(v)	The gender of a person is an example of (a) Categorical and nominal variable (b) Categorical and ordinal variable (c) Discrete Quantitative variable (d) Continuous Quantitative variable	1	CO1	BLT2

NARULA INSTITUTE OF TECHNOLOGY
An Autonomous Institute under MAKAUT

- | | | | | |
|--------|--|---|-----|------|
| (vi) | During training phase, dataset with target class labels is required in | 1 | CO2 | BLT1 |
| | (a) Supervised Learning | | | |
| | (b) Unsupervised Learning | | | |
| | (c) Both Supervised and Unsupervised Learning | | | |
| | (d) None of the above | | | |
| (vii) | The quality of Regression Analysis may be affected by | 1 | CO1 | BLT2 |
| | (a) Collinearity only | | | |
| | (b) Heteroskedasticity only | | | |
| | (c) Both Collinearity and Heteroskedasticity | | | |
| | (d) None of the above | | | |
| (viii) | DBSCAN clustering scheme is one kind of | 1 | CO3 | BLT1 |
| | (a) Hierarchical clustering strategy | | | |
| | (b) Density based clustering strategy | | | |
| | (c) Partitioning clustering strategy | | | |
| | (d) None of the above | | | |
| (ix) | A ReLu activation function for input x may be defined as | 1 | CO5 | BLT2 |
| | (a) $\max(-1, x)$ | | | |
| | (b) $\min(0, x)$ | | | |
| | (c) $\max(1, x)$ | | | |
| | (d) $\max(0, x)$ | | | |
| (x) | Which of the followings is associated with ANN based learning? | 1 | CO5 | BLT1 |
| | (a) Gradient Descent Optimization | | | |
| | (b) Particle Swarm Optimization | | | |
| | (c) Genetic Algorithm Based Optimization | | | |
| | (d) All of the above | | | |
| (xi) | A binary sigmoid function has range of | 1 | CO5 | BLT1 |
| | (a) $(-1, +1)$ | | | |
| | (b) $(-1, 0)$ | | | |
| | (c) $(0, 1)$ | | | |
| | (d) None of the above | | | |
| (xii) | In Random Forest Classification strategy involves | 1 | CO2 | BLT1 |
| | (a) Bayesian Classifier | | | |
| | (b) Single decision tree classifier | | | |
| | (c) LDA | | | |
| | (d) Multiple decision tree classifiers | | | |

NARULA INSTITUTE OF TECHNOLOGY
An Autonomous Institute under MAKAUT

GROUP – B
(Short Answer Type Questions)
(Answer any three of the following) 3 x 5 = 15

SL	Question	Marks	CO	Blooms Taxonomy Level
2.	(i) What are the general steps of deploying classification model?	3	CO1	BLT1
	(ii) Explain Boxplot with respect to the given data: {199, 201, 236, 269, 271, 278, 283, 291, 301, 303, 341}.	2	CO1	BLT2
3.	(i) Compute precision, recall, F-measure for each of the classes – A, B and C in respect of following classification model's outcome.	5	CO3	BLT3

	GoldLabel_A	GoldLabel_B	GoldLabel_C	
Predicted_A	30	20	10	TotalPredicted_A=60
Predicted_B	50	60	10	TotalPredicted_B=120
Predicted_C	20	20	80	TotalPredicted_C=120
	TotalGoldLabel_A=100	TotalGoldLabel_B=100	TotalGoldLabel_C=100	

This is an example confusion matrix for 3 labels: A, B and C

4. (i) Write the algorithm of KNN classification concept. 2 CO3 BLT1
(ii) Apply the idea of KNN classifier to determine the T-Shirt size of a customer having height 161 cms and weight 61 kg (assume k = 5). 3 CO3 BLT3

Ht. (cms)	Wt. (kgs)	Size
158	58	M
158	59	M
158	63	M
160	59	M
160	60	M
163	60	M
163	61	M
160	64	L
163	64	L

Ht (cms)	Wt. (kgs)	Size
165	61	L
165	62	L
165	65	L
168	62	L
168	63	L
168	66	L
170	63	L
170	64	L
170	68	L

5. (i) Apply the concept of regression model for the following dataset to determine the glucose level of a person having age 55. 3 CO2 BLT3

AGE (X)	GLUCOSE LEVEL (Y)
43	99
21	65
25	79

NARULA INSTITUTE OF TECHNOLOGY
An Autonomous Institute under MAKAUT

42	75
57	87
59	81

- | | | | | |
|----|------|---|---|----------|
| | (ii) | How can we evaluate a regression model? | 2 | CO2 BLT5 |
| 6. | (i) | Discuss on maximum-margin hyperplane algorithm of SVM for solving classification problem? | 3 | CO5 BLT2 |
| | (ii) | Justify the utility of kernel trick in SVM for solving classification problem. | 2 | CO5 BLT5 |

GROUP – C
(Long Answer Type Questions)

(Answer any three of the following) 3 x 15 = 45

- | SL | Question | Marks | CO | Blooms Taxonomy Level |
|----|---|-------|-----|-----------------------|
| 7. | (i) What is the working principle of Reinforced Learning in Machine Learning? | 2 | CO1 | BLT1 |
| | (ii) Explain overfitting and underfitting problem along with some remedial measures. | 3 | CO2 | BLT2 |
| | (iii) Develop a decision tree classifier based on the ID3 algorithm with respect to the following dataset and determine whether one should play tennis given outlook=sunny, humidity=high, wind=strong. | 10 | CO3 | BLT3 |

Day	Outlook	Humidity	Wind	Play Tennis
1	Sunny	High	Weak	No
2	Sunny	High	Strong	No
3	Overcast	High	Weak	Yes
4	Rain	High	Weak	Yes
5	Rain	Normal	Weak	Yes
6	Rain	Normal	Strong	No
7	Overcast	Normal	Strong	Yes
8	Sunny	High	Weak	No
9	Sunny	Normal	Weak	Yes
10	Rain	Normal	Weak	Yes
11	Sunny	Normal	Strong	Yes
12	Overcast	High	Strong	Yes
13	Overcast	Normal	Weak	Yes
14	Rain	High	Strong	No

- | | | | | |
|----|-----|--|---|----------|
| 8. | (i) | Trace the iterative steps of K-Medoid algorithm with reference to the following dataset, which is to be divided into two clusters. | 8 | CO4 BLT3 |
|----|-----|--|---|----------|

Id	Weight	Height	Width
1	5	3	10
2	10	15	20
3	15	12	15

NARULA INSTITUTE OF TECHNOLOGY
An Autonomous Institute under MAKAUT

4	24	10	25
5	30	45	30
6	85	70	40
7	71	80	40
8	60	78	50
9	55	52	80
10	80	91	60

- (ii) Is there any advantage of K-Medoid clustering strategy over K-Means clustering strategy? 1 CO4 BLT5

- Apply DBSCAN clustering scheme on following dataset where it is given that the minimum number of points inside the circle around a core-point being at center is 4 and the radius of the circle is 1.9 units. 6 CO4 BLT3

Point	X	Y
P1	3	7
P2	4	6
P3	5	5
P4	6	4
P5	7	3
P6	6	2

Point	X	Y
P7	7	2
P8	8	4
P9	3	3
P10	2	6
P11	3	5
P12	2	4

9. (i) Compare Hierarchical Divisive Clustering strategy with Hierarchical Agglomerative Clustering strategy. 3 CO4 BLT2

- (ii) Apply the concept of Hierarchical Agglomerative Clustering strategy based on the following distance matrix to construct the dendrogram as diagrammatic representation of the entire clustering process. 10 CO4 BLT3

	P1	P2	P3	P4	P5
P1	0				
P2	1.04139	0			
P3	0.59304	0.77369	0		
P4	0.46098	0.61612	0.30232	0	
P5	0.81841	0.32388	0.45222	0.35847	0

- (iii) How Silhouette Coefficient is used to evaluate the quality of a clustering model? 2 CO4 BLT2

10. (i) Propose a supervised learning strategy for retrieving a list of similar image objects for a given query image object based on various image features. 3 CO3 BLT6

NARULA INSTITUTE OF TECHNOLOGY
An Autonomous Institute under MAKAUT

- (ii) Illustrate the concept of Naïve Bayes Classification strategy based on the following dataset and determine whether given fruit is a 'Banana' or 'Orange' or 'Other' when only the 3 features (long, 5 CO3 BLT3

Type	Long	Not Long	Sweet	Not Sweet	Yellow	Not Yellow	Total
Banana	400	100	350	150	450	50	500
Orange	0	300	150	150	300	0	300
Other	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000

sweet and yellow) are known.

- (iii) Apply the principle of apriori algorithm to generate association rules with reference to the following transaction dataset of items where it is assumed that the minimum support value is 2 and confidence is 60%. 7 CO4 BLT3

Transaction Id	Items
1	Hotdogs, Buns, Ketchup
2	Hotdogs, Buns
3	Hotdogs, Chips, Coke
4	Chips, Coke
5	Chips, Ketchup
6	Chips, Coke, Hotdogs

11. (i) How the physiological behaviour of Biological Neuron can be mathematically modeled? 2 CO5 BLT2
- (ii) Discuss on the followings. 4 CO5 BLT2
- (i) Convolutional Neural network
- (ii) Recurrent Neural Network
- (iii) Derive the back propagation formulation model for ANN based classification scheme. 9 CO5 BLT3