June-2022 PG Examinations

USN | | | | | | | | | | |

18MCE13

RV COLLEGE OF ENGINEERING®

(An Autonomous Institution affiliated to VTU, Belagavi)

I Semester Master of Technology (Computer Science and Engineering)

DATA SCIENCE

Time: 03 Hours ***Maximum Marks: 100***

Instructions to candidates:

1. Each unit consists of two questions of 20 marks each.
2. Answer FIVE full questions selecting one from each unit.

UNIT-1

1	a	How is Machine learning different from Data Mining> Explain with the help of an example.	06
	b	Illustrate with an example, 'rules with exceptions' and 'rules with relations'.	10
	c	Paraphrase enumerating a concept space.	04
		OR	
2	a	List and explain properties of bias of the search in machine learning.	10
	b	Discuss over fitting and under fitting scenario in a learning model.	10

UNIT-2

3	a	Briefly explain the major data science project roles and responsibilities.	10
	b	Explain why precision and recall are the major essential quality measures of an information retrieval system. Compute the precision and recall of the model at the threshold values given below:	10

		+	-
Actual	+	3	2
	-	1	4

OR

4	a	With the help of a diagram, explain the stages of a data science project.	12
	b	Explain bias and variance. How can it be overcome in a model.	08

UNIT-3

5	a	Consider the following dataset to answer: i) What is the entropy of this collection of training examples with respect to positive class? ii) What are the information gains of A1 and A2, relative to these training examples? iii) What is the best split (between A1 and A2) according to the Gini index?	

		$Instance$	A_1	A_2	A_3	$Target\ class$
		1	T	T	1.0	$+$
		2	T	T	6.0	$+$
		3	T	F	5.0	$-$
		4	F	F	4.0	$+$
		5	F	T	7.0	$-$
		6	F	T	3.0	$-$
		7	F	F	8.0	$-$
		8	T	F	7.0	$+$
		9	F	T	5.0	$-$	12
	b	Explain the need and criterion for tree pruning.					08
		OR					
6	a	With the help of an example for each, write a note on set, frequent item set, support and confidence.					08
	b	Consider the contingency table-1 and table-2 given below					

	B	$\bar{B}$
A	9	1
$\bar{A}$	1	89

	B	$\bar{B}$
A	89	1
$\bar{A}$	1	9

i) For table-1, compute support and confidence of rules $A \rightarrow B$ and $B \rightarrow A$

ii) For table-2, compute support and confidence of rules $A \rightarrow B$ and $B \rightarrow A$. 08

c Consider the dataset shown below:

$Record$	A	B	C	$class$
1	0	0	0	$+$
2	0	0	1	$-$
3	0	1	1	$-$
4	0	1	1	$-$
5	0	0	1	$+$
6	1	0	1	$+$
7	1	0	1	$-$
8	1	0	1	$-$
9	1	1	1	$+$
10	1	0	1	$+$

Estimate the conditional probabilities for $P(A/+)$, $P(B/+)$, $P(C/+)$, $P(A/-)$, $P(B/-)$, $P(C/-)$ 04

UNIT-4

7	a	Compare and analyze linear regression and Logistic regression in detail with suitable examples.	10
	b	Design Perceptron learning algorithm for linearly separable hyper plane.	10
		OR	
8	a	Discuss bagging, boosting and stacking in detail with an example for each of them.	10
	b	How are Bayesian networks drawn? Explain with the help of an example. Is Bayesian network useful for classification or prediction?	10

UNIT-5

9	a	Why is unsupervised learning considered more challenging than supervised learning?	05
	b	Explain hierarchical clustering algorithm considering the common types of linkages.	10
	c	Why are support vectors considered most important for classification?	05
		OR	
10	a	Explain K-means clustering with an example.	10
	b	Write a note on the following: i) Principle Component analysis ii) Cross validation iii) Stacked Generalization iv) Non linear decision boundaries	10