USN | 1 | S | I | | | | | | | |

MOBILE PHONES PROHIBITED

RCSE31

Siddaganga Institute of Technology, Tumakuru – 572 103

(An Autonomous Institution affiliated to VTU, Belagavi, Approved by AICTE, New Delhi)

Odd Semester B.E. Computer Science & Engg. Makeup Examinations Mar. – Apr. 2021

Foundations of Data Science

Time: 3 Hours **Max. Marks: 100**

Note : Answer any five questions choosing one full question from each unit.

Unit - I

1 a) "Deployment of a data mining solution can be much less technical". Justify. 6

BL:	3	CO:	1	PO:	1	PSO:	3

b) What is data science? Explain the fundamental data mining tasks with suitable example for each. 8

BL:	2	CO:	1	PO:	1	PSO:	3

c) Explain briefly Data Processing and Big Data. 6

BL:	2	CO:	1	PO:	1	PSO:	3

OR

2 a) Give the crisp model of data mining. Explain in detail its stages. 7

BL:	1	CO:	1	PO:	1	PSO:	3

b) Outline and describe any four benefits of data analytic thinking in Business. 7

BL:	2	CO:	1	PO:	1	PSO:	3

c) Explain the difference between Supervised and Unsupervised learning. 6

BL:	2	CO:	1	PO:	1	PSO:	3

Unit – II

3 a) What is probability estimation? Give all the steps to construct the probability estimation tree with smoothed probabilities. 10

BL:	2	CO:	2	PO:	3	PSO:	3

b) Construct the probability estimation tree for the below given training set.

Sl. No.	A	B	C	Class
1	+	*	25	Y
2	+	%	35	N
3	-	*	40	N
4	-	*	35	Y
5	-	*	25	Y
6	+	*	20	N
7	+	%	30	Y

10

BL:	3	CO:	2	PO:	3	PSO:	3

OR

4 a) Construct the probability estimation tree with smoothed probabilities for the following data set.

Sl. No.	House owner	Marital status	Balance	Class: Defaulted
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

10

BL:	3	CO:	2	PO:	3	PSO:	3

Please Turn Over

b) Determine how attribute selection helps in constructing the decision tree. 6

BL:	2	CO:	2	PO:	3	PSO:	3

c) Explain in detail the difference between predictive and descriptive modelling. 4

BL:	2	CO:	2	PO:	3	PSO:	3

Unit – III

5 a) What are Support Vector Machines? With a neat diagram explain in detail the objective function of a Support Vector Machine. 10

BL:	2	CO:	3	PO:	2	PSO:	3

b) For the below given training set develop a linear discriminant model for classification and classify the last Instance. 10

Sl. No.	1	2	3	4	5	6	7
Attr. A	20	24	30	38	43	41	35
Attr. B	50K	50K	30K	100K	120K	90K	85K
Class	Yes	Yes	Yes	No	No	Yes	??

BL:	3	CO:	3	PO:	2	PSO:	3

OR

6 a) Illustrate Repression using mathematical functions. 7

BL:	2	CO:	3	PO:	2	PSO:	3

b) Explain with example how you can get class probability estimation using logistic repression. 5

BL:	2	CO:	3	PO:	2	PSO:	3

c) Develop a linear discriminant function for the following sample data set and predict the class value for the 8^{th} tuple. Also compute the value of f(x). 8

Sl. No.	Age	Balance	Class
1	20	40K	Yes
2	55	110K	No
3	30	30K	Yes
4	25	35K	Yes
5	45	120K	No
6	40	80K	No
7	25	50K	Yes
8	50	95K	??

BL:	3	CO:	3	PO:	2	PSO:	3

Unit – IV

7 a) Discuss on the issues that arise in nearest neighbour method. 8

BL:	2	CO:	4	PO:	3	PSO:	3

b) Differentiate between majority voting and similarity moderated voting. 5

BL:	2	CO:	4	PO:	3	PSO:	3

c) Explain how clustering is carried out around centroids. 7

BL:	2	CO:	4	PO:	3	PSO:	3

OR

8 a) How K-means clustering is done? Explain with an example. 7

BL:	2	CO:	4	PO:	3	PSO:	3

Contd…

b) List few business problems or tasks based on Similarity. 4

BL:	2	CO:	4	PO:	3	PSO:	3

c) Construct a single link dendogram for the given data set. Show all the intermediate distance matrices.

Sl. No.	A	B	C
1	20	32	10
2	25	41	12
3	30	42	13
4	30	41	15
5	25	40	16
6	28	28	18

9

BL:	3	CO:	4	PO:	3	PSO:	3

Unit – V

9 a) How do we evaluate the binary classifier? Explain. 7

BL:	2	CO:	5	PO:	2	PSO:	3

b) Explain how do you measure the term frequency and inverse document frequency with a suitable example. 6

BL:	2	CO:	5	PO:	2	PSO:	3

c) How line prediction and social recommendation are interrelated? 7

BL:	2	CO:	5	PO:	2	PSO:	3

OR

10 a) What are the basic measures for text retrieval? Discuss in detail the text retrieval methods. 8

BL:	2	CO:	5	PO:	2	PSO:	3

b) Define Bag of Words. Give the steps to calculate the term frequency of a word document. 6

BL:	2	CO:	5	PO:	2	PSO:	3

c) Discuss in detail the applications of data mining for the retail industry. 6

BL:	2	CO:	5	PO:	2	PSO:	3