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)

Supplementary Semester B.E. Computer Sc. and Engg. Examinations Sept. 2023

Foundations of Data Science

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

Note : 1. Revealing of Identity in any form in the answer book/graph sheet will be treated as malpractice.
2. Answer any five questions choosing one full question from each unit.

			M	BL	CO	PO	PSO
		Unit - I					
1	a)	What is data science? Explain the fundamental data mining tasks with suitable example for each.	8	2	1		
	b)	List and explain any four data mining tasks.	6	2	1		
	c)	Distinguish between Bigdata 1.0 and Bigdata 2.0.	6	4	1		
		OR					
2	a)	With a neat diagram, explain how data science is placed in the context of various data related process.	8	2	1		
	b)	Illustrate CRISP model for data mining and its stages.	6	2	1		
	c)	Distinguish between i) Machine learning and KDD ii) Predictive and descriptive modeling	6	4	1		
		Unit - II					
3	a)	Demonstrate with example how attribute selection helps in constructing the decision tree.	8	2	2		
	b)	Construct probability estimation tree with smoothed probabilities for the following dataset.					

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

12 3 2

OR

4	a)	What is probability estimation? Give the steps to construct probability estimation tree with smoothed probabilities.	8	2	2		
	b)	Construct a probability estimation tree with probability for the following data set.					

Name	Body temperature	Gives birth	Four legged	Hibernates	Target class
Human	Warm-blooded	Yes	No	No	Yes
Pigeon	Warm-blooded	No	No	No	No
Elephant	Cold-blooded	Yes	Yes	No	Yes
Leopard	Cold-blooded	Yes	No	No	No
Turtle	Cold-blooded	No	Yes	No	No
Penguin	Cold-blooded	No	No	No	No
Eel	Cold-blooded	No	No	No	No
Dolphin	Warm-blooded	Yes	No	No	Yes
Spinyanteater	Warm-blooded	No	Yes	Yes	Yes
Gila monster	Cold-blooded	No	Yes	Yes	No

12 3 2

		Unit - III					
5	a)	Illustrate support vector machines and objectives function with a neat diagram.	6	2	3		
	b)	Describe the following with mathematical functions. i) SVM ii) Regression	6	2	3		

Please Turn Over

c) Write the equation of a linear model and generate a linear discriminant function for the following data and predict the target value for the sixth and seventh tuple?

Sl No.	Age	Income	Class
1	20	40k	Yes
2	25	50k	Yes
3	30	30k	Yes
4	35	110k	No
5	40	120k	No
6	40	80k	?
7	30	115k	?

8 3 3

OR

6 a) Differentiate classification trees with linear classifiers. 6 2 3

b) Interpret a comparison between logistic regression and tree induction. 6 2 3

c) For the given training set fit a linear discriminant model for classification and classify the last instance.

Sl.No.	1	2	3	4	5	6	7	8
Attribute A	20	24	30	38	43	41	35	45
Attribute B	50k	50k	30k	100k	120k	90k	115k	102k
Class	Yes	Yes	Yes	No	No	Yes	No	?

8 3 3

Unit - IV

7 a) Discuss on the issue that arise in nearest neighbor methods. 8 2 4

b) Find the distance between 1 and all others in the following table and identify the k-nearest neighbors.

Customer ID	Gender	Car type	Shirt size	Class
1	M	Family	S	C_o
2	M	Sports	M	C_o
3	M	Sports	M	C_o
4	M	Sports	L	C_o
5	M	Sports	XL	C_o
6	M	Sports	XL	C_o
7	F	Sports	S	C_o
8	F	Sports	S	C_o
9	F	Sports	M	C_o
10	F	Luxury	L	C_o
11	M	Family	L	C_1
12	M	Family	XL	C_1
13	M	Family	M	C_1
14	M	Luxury	XL	C_1
15	F	Luxury	S	C_1
16	F	Luxury	S	C_1
17	F	Luxury	M	C_1
18	F	Luxury	M	C_1
19	F	Luxury	M	C_1
20	F	Luxury	L	C_1

12 3 4

OR

8 a) How k-mean clustering is done? Explain with an example. 8 2 4

b) Consider the following data set, classify Punit based on following methods.
i) Majority voting ii) Similarity moderated voting

Student	Age	Marks	Credits	Placed
Arjun	25	75	8.3	Yes
Bhim	45	23	6.8	No
Charan	30	29	5.6	Yes
Suraj	35	60	7.3	No
Shyam	22	68	4.9	Yes
Punit	42	50	3.8	??

12 3 4

Contd…

Unit - V

9	a)	Explain how link prediction and social recommendation are interrelated?	6	2	5
	b)	Define bag of words. Give the steps to calculate the term frequency of a word document.	6	2	5
	c)	Analyze the basic measures for text retrieval and its methods in detail.	8	4	5
		OR			
10	a)	Outline a short note on: i) Link prediction and social recommendation ii) Evaluating classifiers	10	2	5
	b)	What is meant by association analysis? Compute, i) Support ii) Confidence iii) Lift iv) Leverage if the 10000 transactions analyzed, the data shows that 6000 of the customer transactions include computer games, while 75000 include videos and 4000 include both computer games and videos. For the rule, Boys (X, "Computer games") $\Longrightarrow$ boys (X, "Video").	10	4	5

———