SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)

Department of Computer Engineering
Academic Year 2021-2022
Term Test – I

Course Name: Data Mining and Warehousing **Course Code: DJ19CEC501**
Class: TE (A & B) **Sem:** V
Date: 22/10/2021 **Time**: 11:10 am – 12:10 pm
Maximum Marks: 25

Instructions:

1. Please solve questions in order with clear and dark ink pens
2. Draw figures wherever required
3. Write SAPID on each page top right corner and Sign with Name at the end of each page

Q. No	Questions	Marks
1.	Comment on all the data pre-processing techniques that uses binning. Explain in detail how binning is used in all these techniques?	04
2.	Consider a fashion brand has two stores that sell their products. The brand recorded the number of sales made each month at each store. In the past 12 months, the following was the sales data. Store 1: 350, 460, 20, 160, 580, 250, 210, 120, 200, 510, 290, 380 Store 2: 520, 180, 260, 380, 80, 500, 630, 420, 210, 70, 440, 140 a. Draw box plots for each sales store. b. Give your analysis for the sales of each store.	 02 01
3 a.	Differentiate between bagging and boosting.	02
3 b.	Explain with an example the process of learning a rule in rule-based classification.	02
3.	The following table consists of training data from an employee database. The data have been generalized. For example, "31 ... 35" for age represents the age range of 31 to 35. For a given row entry, count represents the number of data tuples having the values for department, status, age, and salary given in that row. Let status be the class label attribute. Construct a decision tree using ID3 technique. Draw the final tree.	06

department	*status*	*age*	*salary*
sales	senior	31...35	46K...50K
sales	junior	26...30	26K...30K
sales	junior	31...35	31K...35K
systems	junior	21...25	46K...50K
systems	senior	31...35	66K...70K
systems	junior	26...30	46K...50K
systems	senior	41...45	66K...70K
marketing	senior	36...40	46K...50K
marketing	junior	31...35	41K...45K
secretary	senior	46...50	36K...40K
secretary	junior	26...30	26K...30K

4 a.	Discuss the advantages and disadvantages of k-means clustering method.	02
4 b.	Apply k-mediods for the following distance matrix for 2 clusters and find the clusters. Justify your choice of final cluster	06

Item	A	B	C	D
A	0	1	4	5
B	1	0	2	6
C	4	2	0	3
D	5	6	3	0