SRM Institute of Science and Technology
College of Engineering and Technology
School of Computing

SRM Nagar, Kattankulathur – 603203, Chengalpattu District, Tamil Nadu
Academic Year: 2022-23 (Odd)

Test: CLA-T1
Course Code & Title: 18CSE355T - Data Mining and Analytics
Year & Sem: III Year / V Sem

Date: 08.09.2022
Duration: 1 Hour
Max. Marks: 25 Marks
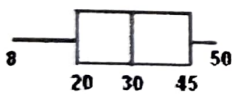
Course Articulation Matrix: *(to be placed)*

S.No.	Course Outcome	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
1	CO1	2	2	1									

Part – A (15 x 1 = 15 Marks)

Instructions: 1. Answer all questions.
2. The duration for answering the part A is 20 minutes (this sheet will be collected after 20 minutes)
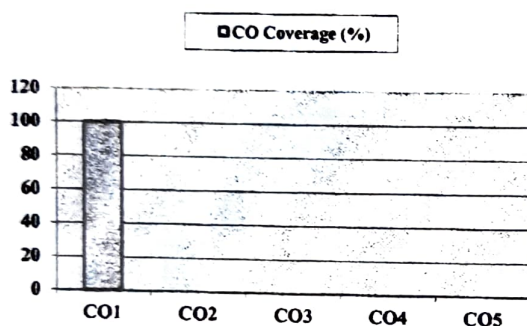
Q. No	Question	Mark	BL	CO	PO	PI Code
1.	The output of KDD is _____ A. Data B. Information C. Query D. Useful Information	1	1	1	1	1.6.1
2.	The data objects that do not comply with the general behaviour or model is called _____ A. Outliers B. Prediction C. Evolution analysis D. Classification	1	1	1	1	1.6.1
3.	"Handling complex types of data" in data mining algorithms issues comes under the category _____ A. Diverse Data type B. Descriptive data type C. Mining Methodology and user interaction issue D. Performance issue	1	1	1	1	1.6.1
4.	Which of the following analyse class-labelled (training) data sets? i. Clustering, ii. Classification, iii. regression A. i only B. ii only C. ii and iii D. i, ii and iii	1	1	1	1	1.6.1
5.	"Privacy-preserving data mining" - This issue falls under which category of issues in data mining? A. Mining Methodology B. Performance Issues C. Data mining and Society D. User Interaction	1	1	1	1	1.6.1
6.	The values of a/an _____ are measured in fixed and equal units. A. Ratio-Scaled Attribute B. Time-Scaled Attribute C. Interval-Scaled attribute D. Point-Scaled Attribute	1	1	1	1	1.6.1

#	Question					
7.	Identify which is the ratio scaled attribute among the given attributes A. Rank of officers B. Temperature in Fahrenheit C. Temperature in Kelvin D. Grades of students	1	1	1	1	1.6.1
8.	What is the term for the median of the lower half of the data? A. Q1 B. Q3 C. IQR D. Maximum	1	1	1	1	1.6.1
9.	The analysis tools pre-compute the summaries of the huge amount of data for what purpose? A. In order to maintain consistency B. For authentication C. For data access D. To obtain the query response	1	1	1	1	1.6.1
10.	What is a median? A. Difference between higher half and lower half of the data set B. Mean of the highest and lowest number in a data sample C. Value separating higher half from the lower half of a data sample D. Difference between the highest and lowest number.	1	1	1	1	1.6.1
11.	Given the mean of the dataset is 13 and the standard deviation is 5. Using z-score normalization to transform the value 8, gives a new value of A. 0 B. 1 C. -1 D. -1.5	1	2	1	2	2.6.3
12.	Based on the box and whisker plot below, what is the IQR of the data? A. 30 B. 25 C. 20 D.15	1	2	1	2	2.6.3
13.	What is the Median of the following data sample? 2, 7, 4, 8, 9, 10, 6, 12, 13 A. 8 B. 11 C. 9 D. 10	1	2	1	2	2.6.3
14.	Use these method to normalize the following group of data: 200, 300, 400, 600 and 1000. Use min-max normalization by setting min=0, and max=1 for the value 600. A. 0.4 B. 0.5 C. 0.6 D. 0.7	1	2	1	2	2.6.3
15.	A survey was given to a random sample of 20 college students. They were asked, "how many textbooks do you own?". Their responses, were: 0, 0, 2, 5, 8, 8, 8, 9, 9, 10, 10, 10, 11, 12, 12, 12, 14, 15, 20, and 25. Which among the following is TRUE? A. Q1=20 B. Q3=0 C. Median=8 D. The value 20 is an outlier	1	2	1	2	2.6.3

	PART B (1 X 10 = 10 Marks) Instruction: Answer either A or B					
16 A.	i. Can all kinds of data be mined? Discuss about what kinds of data can be mined.	6	2	1	1	1.6.1
	ii. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 36, 40, 52, 70. Find out the Five number summary for the above data.	4	3	1	2	2.6.3
16 B.	i. Justify how attribute subset selection helps to make patterns easier to understand.	6	2	1	1	1.6.1
	ii. Normalize the following group of data: 1000,2000,3000,9000 using min-max normalization by setting min:0 and max:1	4	3	1	2	2.6.3

*Program Indicators are available separately for Computer Science and Engineering in AICTE examination reforms policy.

Course Outcome (CO) and Bloom's level (BL) Coverage in Questions

Approved by the Audit Professor/Course Coordinator