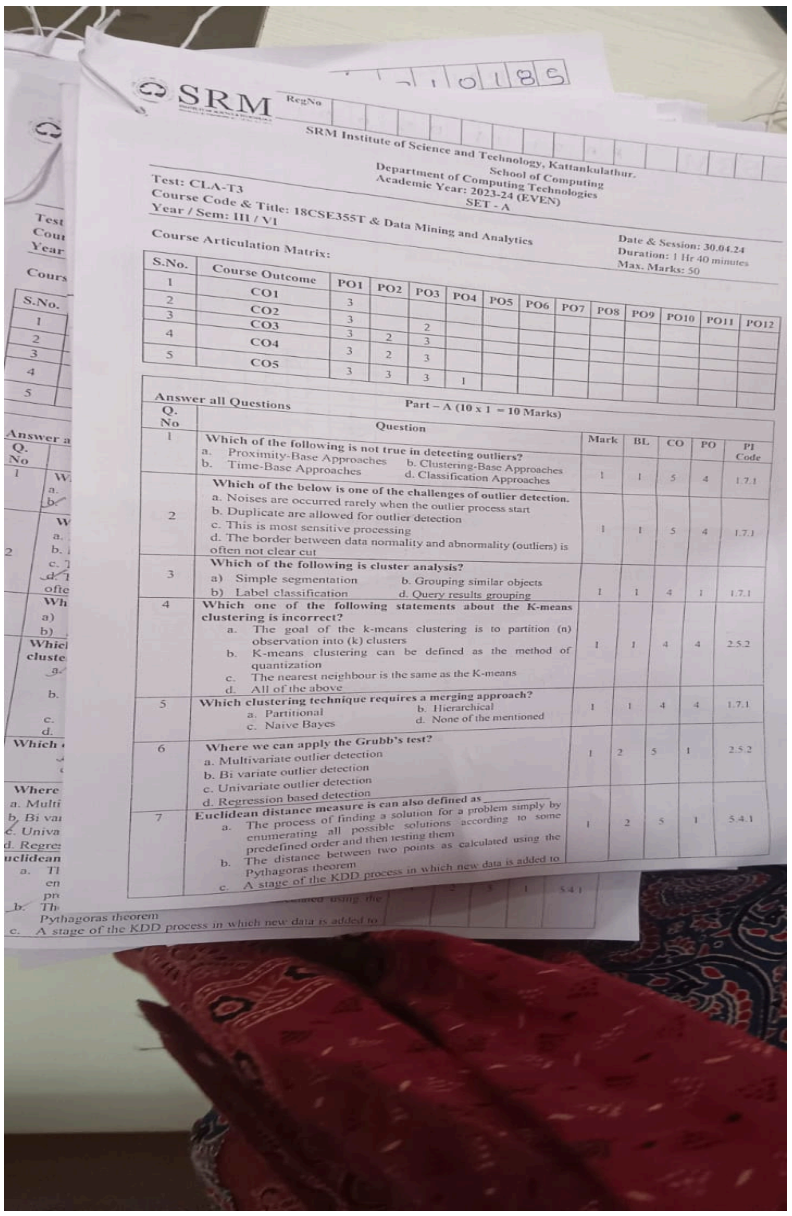

SRM

Reg No 7 1 1 0 1 8 5

SRM Institute of Science and Technology, Kattankulathur.
School of Computing
Department of Computing Technologies
Academic Year: 2023-24 (EVEN)
SET - A

Test: CLA-T3
Course Code & Title: 18CSE355T & Data Mining and Analytics
Year / Sem: III / VI

Date & Session: 30.04.24
Duration: 1 Hr 40 minutes
Max. Marks: 50

Course Articulation Matrix:

S.No.	Course Outcome	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
1	CO1	3											
2	CO2	3		2									
3	CO3	3	2	3									
4	CO4	3	2	3									
5	CO5	3	3	3	1								

Answer all Questions

Part – A (10 x 1 = 10 Marks)

Q. No	Question	Mark	BL	CO	PO	PI Code
1	**Which of the following is not true in detecting outliers?** a. Proximity-Base Approaches b. Clustering-Base Approaches b. Time-Base Approaches d. Classification Approaches	1	1	5	4	1.7.1
2	**Which of the below is one of the challenges of outlier detection.** a. Noises are occurred rarely when the outlier process start b. Duplicate are allowed for outlier detection c. This is most sensitive processing d. The border between data normality and abnormality (outliers) is often not clear cut	1	1	5	4	1.7.1
3	**Which of the following is cluster analysis?** a) Simple segmentation b. Grouping similar objects b) Label classification d. Query results grouping	1	1	4	1	1.7.1
4	**Which one of the following statements about the K-means clustering is incorrect?** a. The goal of the k-means clustering is to partition (n) observation into (k) clusters b. K-means clustering can be defined as the method of quantization c. The nearest neighbour is the same as the K-means d. All of the above	1	1	4	4	2.5.2
5	**Which clustering technique requires a merging approach?** a. Partitional b. Hierarchical c. Naive Bayes d. None of the mentioned	1	1	4	4	1.7.1
6	**Where we can apply the Grubb's test?** a. Multivariate outlier detection b. Bi variate outlier detection c. Univariate outlier detection d. Regression based detection	1	2	5	1	2.5.2
7	**Euclidean distance measure is can also defined as** a. The process of finding a solution for a problem simply by enumerating all possible solutions according to some predefined order and then testing them b. The distance between two points as calculated using the Pythagoras theorem c. A stage of the KDD process in which new data is added to	1	2	5	1	5.4.1

SRM

RegNo

	the existing selection. d. It is a kind of process of executing implicit, previously unknown and potentially useful information from data					
8	**The analysis performed to uncover the interesting statistical correlation between associated -attributes value pairs are known as the _____.** a. Mining of association b. Mining of correlation b. Mining of clusters d. Mining of Prediction	1	2	4	4	2.5.2
9	**The K means clustering algorithm fails to give good results in _** a. When the dataset contains outliers. b. When the data points follow a non-convex shape. c. When the data points follow a convex shape. d. Both a and b	1	2	4	4	2.5.2
10	**What is the time complexity of K means clustering algorithm?** a. O(K) b. O(n-1) c. O(n) d. O(nkt)	1	2	5	2	2.5.2
	Part – B (5 x 4 = 20 Marks)					
	Answer five Questions					
11	Write K-Medoids clustering algorithm with an example.	4	3	4	2	2.5.2
12	Differentiate between AGNES and DIANA algorithms.	4	2	4	1	2.5.2
13	Discuss about STING method from grid based clustering algorithm.	4	3	4	2	2.5.2
14	Explain different types of outlier with an example.	4	2	5	4	2.5.2
15.	Explain Data Mining for Financial data analysis	4	3	5	2	2.5.2
16.	Differentiate Supervised and Unsupervised Learning	4	2	4	1	2.5.2
17.	Explain the different Challenges of Outlier Detection	4	3	5	2	2.5.2
	Part – C (2 x 10 = 20 Marks)					
18	Consider the Following data points to compute Cluster Values when number of clusters is three using Partitioning Clustering Algorithm with mean as a distance metrics : K= {X1(2,10), X2(2,5), X3(8,4), X4(5,8), X5(7,5), X6(6.4), X7(1,2), X8(4,9)}.	10	3	4	2	2.5.2
	OR					
19	What is a DBSCAN? Apply DBSCAN algorithm to the given data points to create the cluster with minpts = 4, epsilon = 1.9 and p1(3,7),p2(4,6),p3(5,5),p4(6,4),p5(7,3),p6(6,2),p7(7,2),p8(8,4), p9(3,3),p10(2,6),p11(3,5),p12(2,4).	10	3	4	2	2.5.2
20	Discuss about attributes of healthcare recommendation system using Data mining approach with example.	10	2	5	4	2.7.1
	OR					
21	Interpret the supervised method for detecting the outlier.	10	2	5	4	1.7.1

*Performance 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

SRM

Reg. No.

SRM Institute of Science and Technology, Kattankulathur.
School of Computing Department of
Computing Technologies
Academic Year: 2023-24 (EVEN)SET - B

Test: CLA-T3
Course Code & Title: 18CSE355T & Data Mining and Analytics
Year / Sem: III / VI

Date & Session: 30.04.24
Duration: 1 Hr 40 minutes
Max. Marks: 50

Course Articulation Matrix:

S.No	Course Outcome	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO
1	CO1	3											
2	CO2	3		2									
3	CO3	3	2	3									
4	CO4	3	2	3									
5	CO5	3	3	3	1								

Part – A (10 x 1 = 10 Marks)

All questions. The duration for answering the part A is 15 minutes (MCQ Answersheet will be collected after 15 minutes)

	Question	Marks	BL	CO	PO
1	**Which of the below one is not an typicalrequirements of clustering in data mining.** a. Scalability b. Ability to deal with different types of attributes c. Ability to deal with noisy data d. Decremental clustering and insensitivity to inputorder	1	1	4	1
2	**Given a set of n objects, a partitioning methodconstructs k partitions of the data, where each partition represents a cluster and k ≤ n. How manygroups in this conditions?** a. k groups b. n groups c. Kn d. n^K	1	2	4	2
3	**Which of the below method is used for the bottom-up approach.** a. K mean method b. Agglomerative method c. Divisive method d. Partitioning method	1	1	4	1
4.	**Which one of the following can be defined as the data object which does not comply with the general behaviour(or the model of available data)?** a. Evaluation Analysis c. Outliner Analysis b. Classification d. Prediction	1	1	5	2
5.	**Let p1=(1,2) and p2=(3,5) represent two objects,what will be the Euclidean distance?** a) 5 b) 3.61 c) 6.31 d) 2	1	2	4	1
6.	**What is the time complexity of K means clusteringalgorithm.** a. O(K) b. O(n-1) c. O(n) d. O(nkt)	1	2	4	2
7.	**-------- uses the notions of clustering feature tosummarize a cluster,** and clustering feature tree (CF-tree) to represent a cluster hierarchy. a. BIRCH b. DBSCAN c. STING d. CLIQUE	1	1	4	1
8	**Which will play the major role in proximity-basedoutlier detection?** a. Density of the neighbourhood b. Mean vector of the neighbourhood c. Radius of the neighbourhood d Threshold of the neighbourhood	1	1	5	1
9	**What method is used to find the intrusion detection in clustering-based outlier detection?** a. Bootstrap b. Angle-based outlier c. Anomalies detection d. CLIQUE	1	2	5	1
10	**In which case objects are labelled as "normal" or"outlier" are not available?** a. Supervised method b. Unsupervised method c. Semi-supervised method d. Hybrid method	1	2	5	1

	Part – B (5 x 4 = 20 Marks) Answer any Five Questions					
11	Write the algorithm for centroid based clustering algorithm.	4	3	4	3	8.4.1
12	Discuss the basic characteristic of clustering methods	4	3	4	3	1.7.1
13	Discuss the outlier detection using histogram-based approach with example	4	3	4	4	1.7.1
14	Explain the various steps involved in statistical based outlier detection with suitable diagram	4	2	5	4	2.6.4
15	Solve the single link technique using below table	4	3	5	2	2.6.4
16	Explain briefly about semi supervised learning tchique for outlier detection	4	2	5	2	2.6.4
17	Explain how data mining used in science and engeering	4	2	5	2	2.6.4
	Part – C (2 x 10 = 20 Marks)					
18	(i)Find the Euclidean distance, Manhattan distance, for below table (ii)Explain the jaccard coefficient with formula	7+3	3	4	2	1.7.1
	OR					
19	Explain the density-based clustering method withpseudo code, suitable diagram with formulas.	10	2	4	1	8.4.1
20	A city's average temperature values in July in the last 10years are, in value-ascending order. 24.0°C, 28.9°C, 28.9°C, 29.0°C, 29.1°C, 29.1°C, 29.2°C, 29.2°C, 29.3°C and 29.4°C. Let's assume that the average temperature follows a normal distribution, which is determined by two parameters: the mean, μ, and the standard deviation, σ.Use the maximum likelihood method to estimate the parameter μ and σ.	10	3	5	2	8.4.1
	OR					
21	How data mining involved in the below application.(i)Data Mining for Financial data analysis (ii) Data Mining for Retail and TelecommunicationIndustries (iii) Data Mining in Science and Engineering (iv) Data Mining for Intrusion Detection and Prevention(v)Data Mining and Recommender Systems	10	2	5	4	8.4.1

	X1	Y1			
P1	1.5	1.7			
P2	2	1.9			
P3	1.6	1.8			
P4	1.2	1.5			
P5	1.5	1.0			
P6	2	2.9			

*Performance 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

CO Coverage in %

BL Coverage in %