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

Batch - 2

DEPARTMENT OF COMPUTING TECHNOLOGIES
SRM Nagar, Kattankulathur – 603203, Chengalpattu District, Tamilnadu

Academic Year: 2022-2023(ODD)

Test: CLAT-3 **Date:** 16.11.2022
Course Code & Title: 18CSE355T - Data Mining and Analytics **Duration:** 2 Periods
Year & Sem: III Year & 05th Semester **Max. Marks:** 50 Marks

Course Articulation Matrix:

S. No.	Course Outcome	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
1	CO4	3							3				
2	CO5		3						3				
3	CO6		2						2				

Part – A
(10 x 1 = 10 Marks)
Answer all questions. The duration for answering the part A is 20 minutes (MCQ Answer sheet will be collected after 20 minutes)

Q. No	Question	Marks	BL	CO	PO	PI Code
1	Which is conclusively produced by Hierarchical Clustering? a) Final estimation of cluster centroid b) Tree showing how nearby things are to each other c) Assignment of each point to clusters d) Assignment of each centroid to clusters	1	1	4	1	1.7.1
2	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	1.7.1
3	Which of the following is cluster analysis? a) Simple segmentation b) Grouping similar objects c) Label classification d) Query results grouping	1	1	4	1	1.7.1
4	_____ method removes sparse clusters as outliers and groups dense clusters in to larger ones. a) DIANA b) BIRCH	1	2	4	1	1.7.1

	c) STING d) DBSCAN					
5	Which is not part of the categories of clustering methods? a) Hierarchical methods b) Density based methods c) Portioning methods d) Rule-based methods	1	1	4	1	1.7.1
6	What are the different ways to classify an Intrusion detection System? a) Zone based b) Host & Network based c) Network & Zone based d) Level based	1	1	5	2	2.7.1
7	Find the outlier in the given data set below. 16, 14, 3, 12, 15, 17, 22, 15, 52 a) 22 b) 12 c) 52 d) 3	1	2	5	2	2.7.1
8	The learning algorithms that can deal with both minimal labelled dataset and large unlabelled dataset together is called _____. a) Supervised b) Unsupervised c) Semi supervised d) Reinforcement	1	1	5	2	2.7.1
9	In customer relationship management, we can detect outlier customers using _____. a) Data sparsity b) Contextual outlier detection c) Collective outlier detection d) Conventional Outlier Detection	1	1	6	2	2.7.1
10	Internet search engine are tasks related to the area of _____. a) Information retrieval b) Information storage c) Information cluster d) Information visualization	1	1	6	8	8.4.2

	Part – B **(4 x 5 = 20 Marks)** **Answer any 4 Questions**					
11	Outline the K-Medoid Clustering Method.	5	2	4	1	1.7.1
12	List the application of cluster analysis.	5	2	4	1	1.7.1
13	How to Measuring Clustering Quality? Explain.	5	2	4	8	8.4.1
14	Write about Generalized linear model and Mixed-effect model.	5	2	5	2	2.6.4
15	Explain in detail how data mining used for Retail and Telecommunication Industries.	5	2	6	2	2.6.4
	Part – C **(2 x 10 = 20 Marks)**					
16	Explain the K-Means algorithm in detail and Apply the K-means algorithm for the following five points (with (x, y) representing locations) into two clusters: A1(3, 10), A2(7, 5), A3(10, 4), A4(5, 9), A5(8, 5). Initial cluster centers are: A1(3, 10), and A4(5, 9)	10	3	4	1	1.7.1
	[OR]					
17	Write about hierarchical clustering methods in detail.	10	3	4	8	8.4.1
18	Interpret the supervised method for detecting the outlier.	10	3	5	2	2.4.1
	[OR]					
19	Explain in detail, how data mining algorithms can be used for Intrusion detection and prevention.	10	3	5	8	8.4.1

***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