BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: MTECH (COGNIZANT) **SEMESTER : I**
BRANCH: CS **SESSION:MO/2022**

SUBJECT: CS538 UNSUPERVISED LEARNING

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
***4*. Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO
Q.1(a)	Briefly outline how to compute the dissimilarity between objects described by the following types of variables: (a) Numerical (interval-scaled) variables (b) Asymmetric binary variables (c) Categorical variables (d) Ratio-scaled variables (e) Nonmetric vector objects	[5]	1,1,1
Q.1(b)	Briefly describe the following approaches to clustering: grid-based methods, model-based methods, methods for high-dimensional data, and constraint-based methods. Give examples in each case.	[5]	1,1,1
Q.2(a)	Suppose that the data mining task is to cluster the following eight points (with (x, y) representing location) into three clusters. A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9). The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. Use the k-means algorithm to show only. (a) The three cluster centers after the first round of execution and (b) The final three clusters	[5]	2,4,2
Q.2(b)	Clustering has been popularly recognized as an important data mining task with broad applications. Give one application example for each of the following cases: (i). An application that takes clustering as a major data mining function (ii). An application that takes clustering as a preprocessing tool for data preparation for other data mining tasks.	[5]	2,2,2
Q.3(a)	Use a diagram to illustrate how, for a constant MinPts value, density-based clusters with respect to a higher density (i.e., a lower value for Є, the neighborhood radius) are completely contained in density-connected sets obtained with respect to a lower density.	[5]	3,5,3
Q.3(b)	Why is it that BIRCH encounters difficulties in finding clusters of arbitrary shape but OPTICS does not? Can you propose some modifications to BIRCH to help it find clusters of arbitrary shape?	[5]	3,6,3
Q.4(a)	Consider the contingency table for the clustering in figure-a, Compute the different pair-wise measures for clustering evaluation.	[5]	4,4,4
Q.4(b)	Compares the different contingency-based measures Jaccard, Rand, and FM on the two clustering in figure-a, b.	[5]	4,1,4

Figure-a.

	Iris-setosa	Iris-versicolor	Iris-virginica
	T1	T2	T3
C1	0	47	14
C2	50	0	0
C3	0	3	36

PTO

Q.4(b) Figure-b.

	Iris-setosa	Iris-versicolor	Iris-virginica
	T1	T2	T3
C1	0	47	14
C2	50	0	0
C3	0	3	36

Q.5(a) Why is outlier mining important? Briefly describe the different approaches behind statistical-based outlier detection, distanced-based outlier detection, density-based local outlier detection, and deviation-based outlier detection. [5] 5,4,5

Q.5(b) Explain briefly the Local outlier factor (LOF). [5] 5,2,5

::::::20/03/2023::::::M