Birla Institute of Technology & Science, Pilani

Work Integrated Learning Programmes Division
Comprehensive Exam (Sample)

Course Number	:	PCAM ZC221	No. of Pages: 1
Course Title	:	Unsupervised Learning and Association Rule Mining	No. of Questions: 5
Type of Exam	:	Closed Book	
Weightage	:	40 %	Duration: 3 hours
Date of Exam	:	__/__/___	Session : FN

1 a. Explain the working of the Mini-Batch K-Means algorithm in the incoming two points in the given order. Given : Cluster 1 = {12, 50, 4, 20} Cluster 2 ={32, 15, 2}
Incoming Data : {10, 5}

b. "Density connectedness is symmetric but Density reachability is asymmetric." Justify the statement with proper explanation using example. 6m

2 Cluster the following points using K-Means algorithm. (1, 1), (1.5, 2.5), (3, 4), (6, 7), (3.5, 5) and (4.5, 5). Let the initial centroids be (1, 1) and (3, 4). Use Manhattan distance as distance measure and mean of cluster points as cluster centroid. Perform two iterations. 6m

3 Consider the below clustered 1-D points, calculate the following cluster validity index and interpret the suitability of the data validation.
Cluster 1 : 47, 49, 42, 41
Cluster 2 : 39, 43, 40, 36
Cluster 3 : 7, 6, 15 8m

a) David – Bouldin Index
b) Dunn Index

4 Assume outlier score of an object is inverse of density around an object. Using this definition, determine the outlier for the objects specified in the following distance matrix. Consider 3 nearest neighbors for the density determination. 10m

	A	B	C	D	E
A	0	1	4	5	7
B	1	0	2	6	8
C	4	2	0	3	4
D	5	6	3	0	4
E	7	8	4	4	0

5 For the market basket transactions given in the following table. Find all the frequent item sets using 10m

a) Apriori algorithm.
b) FP-Tree Growth

Transaction ID	Items Bought
T1	A,B,C
T2	A,B,C,D,E
T3	A,C,D
T4	A,C,D,E
T5	A,B,C,D