

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST-3 EXAMINATION- May, 2023

B.Tech. (BT/BI) VI Semester

COURSE CODE: 18B1WBI632 (3)

MAX. MARKS: 35

COURSE NAME: Dataware Housing and Mining for Bioinformatics

COURSE INSTRUCTORS: Dr. Ekta Gandotra

MAX. TIME: 2 Hrs.

Note: All questions are compulsory. Marks are indicated against each question in square brackets.

- Q1. a. Suppose the data mining task is to cluster the following eight data points (with (x, y) representing location) into three clusters. [5] CO6
 A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7,5), B3(6, 4), C1(1, 2), C2(4, 9).
 Use k-Means algorithm to find the three cluster centers after the second iteration. Take A1, B1 and C1 as the centre for each cluster respectively. Use Euclidean distance as distance function.
- b. What is Silhouette Coefficient? How its value is interpreted to evaluate the quality of a clustering model? [3] CO6
- Q2. a. Find the frequent item sets and generate the association rules for the following dataset [5] CO5
 using Apriori algorithm. Take minimum support=2 and minimum confidence = 50%.

TID	Itemsets
T1	A, B
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C
T7	A, C
T8	A, B, C, E
T9	A, B, C

- b. Explain transaction reduction method to improve the efficiency of Apriori algorithm using an example. [3] CO5
- Q3. a. Consider the following distance matrix. Show the step-wise results of agglomerative hierarchical clustering with single link by drawing a dendrogram. [5] CO6

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.70	0.45	0	
F	0.34	0.61	0.93	0.20	0.67	0

b. Explain the input parameters given to the DBSCAN algorithm. Discuss the type of points we get after applying this algorithm to a particular dataset? [4] CO6

Q4. a. Consider the following dataset of training examples: [3] CO4

A	B	Class Label
T	T	Yes
T	T	Yes
T	F	No
F	F	Yes
F	T	No
F	T	No

Find the information gain of attribute **B** relative to these training examples?

b. Suppose you build a model which shows a training accuracy of 98% and a test accuracy of 58%. What could be the possible reasons for the gap between these accuracies? Suggest a method to overcome this problem. [3] CO3

c. Explain Adaboost algorithm with the help of an example. [4] CO4