COURSE CODE: 10B22CI622   MAX. MARKS: 25

COURSE NAME: Data Mining

COURSE CREDITS: 4   MAX. TIME: 90 Min.

Note: All questions are compulsory.

Q1. CO-2 (a) Use the methods below to normalize the following group of data: [5]
200, 300, 400, 600, 1000
   (i)   min-max normalization by setting min = 0 and max = 1
   (ii)   z-score normalization

Q2. CO-3 (a) Suppose a market shopping data warehouse consists of four dimensions: customer, [3]
date, product, and store, and two measures: count, and avg sales, where avg sales store the
real sales in dollar.
(i) Enumerate three classes of schemas that are popularly used for modeling data
warehouses.
(ii) Draw a schema diagram for the above data warehouse using one of the schema classes
listed in (i).
(b) Write the difference between OLAP and OLTP under the following headings. [2]
(i) DB Design (ii) Data (iii) Usage (iv) Access

Q3. CO-4 (a) Give mathematical formulation of anti-monotone property of support also explain [2]
Apriori principle.

(b) Suppose you are executing Apriori Algorithm for frequent pattern generation and you [2]
obtained L3={abc, abd, acd, ace, bcd} show properly self joining step and pruning step to
obtain C4.

Q4. CO-4 Explain ECLAT Algorithm for Frequent Item set generation. A database has four transactions. [6]
Let *min sup* = 50%. Find all frequent item sets using ECLAT and FP-Growth Algorithms
respectively. Compare the efficiency of the two mining processes

TID	Items_bought
T10	{I1, I3, I4}
T20	{I2, I3, I4}
T30	{I1, I2,I3, I4}
T40	{I2, I4}

Q.5 CO-5 Suppose that the data mining task is to cluster points (with (x, y) representing location) [5]
into three clusters, where the points are **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 i) The three cluster centers after the first round of execution.
ii) The final three clusters