

Total No. of printed pages = 4

63/2 (SEM-3) CSIT 3:3

2022

(Held in 2023)

CSIT

(Theory Paper)

Paper Code : CSIT-3:3

(Data Mining and Warehouse)

Full Marks – 80

Pass Marks – 32

Time – Three hours

**The figures in the margin indicate full marks
for the questions.**

1. Answer the following questions : $2 \times 10 = 20$
- (a) Explain the techniques of data mining.
 - (b) Why we need data transformation ?
 - (c) Define Concept hierarchy.
 - (d) What do you mean by dimension and fact table ?
 - (e) What is the difference between agglomerative and divisive hierarchical clustering ?

[Turn over

- (f) Could you list a few area in which data mining can be applied ?
- (g) Differentiate between Supervised learning and Unsupervised learning.
- (h) What is lattice of cuboids ?
- (i) List out the types of metadata.
- (j) Define Association Rule Mining.

2. Answer any *five* of the following questions :
6×5=30

- (a) Give a brief idea about data processing technologies and difference between them.
- (b) Write a note on Architecture of data warehouse.
- (c) Explain the algorithm for finding frequent item sets using candidate generation.
- (d) How k-Mean clustering method differs from k-Medoids clustering method ? Explain with example.
- (e) Explain Knowledge Discovery process with detail.
- (f) What is the importance of data marts in data warehouse ?

- (g) What are the various types of OLAP Servers ? Explain.
- (h) 'Data preprocessing is necessary before data mining process.' Justify your answer.

3. Answer any *three* of the following questions :
10×3=30

- (a) What is data cube in multidimension data model ? Explain with example. With necessary diagrams and examples of data cubes explain various OLAP operations.
- (b) Describe briefly about the schemas in data warehouse and also describe the syntax of the schemas with example.
- (c) Describe briefly about the different data types in cluster analysis.
- (d) Find the frequent item sets and strong association rules for the following transactional database table using Apriori algorithm. Consider the thresholds as support = 30% and confidence = 40% :

TID	Item IDs
1	11, 12, 13, 15
2	12, 15, 17, 19
3	11, 13, 15, 17

TID	Item IDs
4	12, 14, 16, 18
5	11, 12, 13, 14
6	12, 13, 14, 15
7	13, 14, 15, 16
8	14, 15, 16, 17
9	15, 16, 17, 18, 19
10	19, 11, 12, 15
11	18, 12, 19, 17
12	15, 16, 13, 12