

**M.Sc. Semester-IV Examination 2017**  
**Computer Science**  
**Course : MCSO-48**  
**(Data Mining and Warehousing)**

**Time : 3 Hours**

**Full Marks : 40**

Questions are of value as indicated in the margin

Attempt Question No.1 and **any three** from the rest

1. (a) Briefly explain three basic techniques for data mining with implication of each of those.  
 (b) Draw an itemset lattice with five items and explain the Apriori property using the lattice.  
 (c) What are Precision and Recall? Explain their utilities.  
 (d) Compare supervised and semi supervised classification. 3+3+2+2=10
2. (a) What do you mean by data warehousing? Give an architecture of it and explain its role in association to data mining. What is data mart? 2+2+1+1=6  
 (b) What is multidimensional data? Explain with an example and hence build a data cube from the example. 1+3=4
3. How can decision tree perform as a good classifier? Explain with an example. Write down an algorithm for building decision tree. What are Entropy and Gini Index? Illustrate their role in splitting using your example of decision tree. 3+3+4=10
4. A transaction database is given in the form :

Tid	Items present
1	b d e
2	a c d f
3	d b c
4	a b e
5	a b c
6	e b d
7	a b c f
8	a b c d
9	a b c
10	a b

- (a) With support threshold = 3, build FP-tree for the database and hence obtain frequent itemsets with e from the conditional FP-tree for the item e.
- (b) Obtain Transaction Prefix (TP) tree with support Threshold = 4. 4+3+3=10

(2)

5. (a) Describe the basic principles of DBSCAN and K- means clustering methods.

Discuss their merits and demerits

(2+2)+2=6

(b) Write down the algorithm for DBSCAN. Illustrate density connectivity with example. 3+1=4

6. Write short notes on **any two** :

5+5=10

(a) Apriori Algorithm and scope of its improvement.

(b) Support vector machine.

(c) Nondeterministic classification and clustering.

---