

Third Semester M.Tech. Degree Examination, Dec. 2013/Jan. 2014

Data Mining and Warehousing

Time: 3 hrs.

Max. Marks:100

Note: Answer any FIVE full questions.

- 1 a. Define data warehouse. Explain the architecture of data warehouse, with a neat block diagram. (10 Marks)
- b. Discuss data integration. (05 Marks)
- c. Normalize the following group of data using Z – score normalization : 200, 300, 400, 600, 1000. (05 Marks)
- 2 a. List and describe the five primitives for specifying a data mining task. (10 Marks)
- b. Write DMQL syntax for defining each of the five primitives, with suitable examples. (10 Marks)
- 3 a. Generate strong association rules for the given database assuming min – sup = 60% and min – conf = 80% using Apriori algorithm. (10 Marks)

TID	Items bought
T ₁₀₀	{M, O, N, K, E, Y}
T ₂₀₀	{D, O, N, K, E, Y}
T ₃₀₀	{M, A, K, E}
T ₄₀₀	{M, U, C, K, Y}
T ₅₀₀	{C, O, O, K, I, E}

- b. Briefly explain the approaches to mining multilevel association rules. (10 Marks)
- 4 a. Explain the basic algorithm for inducing a decision tree from training samples, with suitable example. (10 Marks)
- b. Explain any two classification methods, in detail. (10 Marks)
- 5 a. The following table shows the salary of college graduates.

X(years of experience)	Y(salary in \$ 1000 s)
3	30
8	57
9	64
13	72
3	36
6	43
11	59
21	90
1	20
16	83

- i) Use the method of least squares to find an equation for the prediction of a college graduate's salary based on the years experience
- ii) Predict the salary of a graduate with 10 years of experience. (10 Marks)
- b. Calculate simple matching coefficient and Jaccard coefficient for the following two binary vectors.
 - i) $X = (1, 0, 1, 0, 0, 0)$ $Y = (1, 0, 1, 0, 1, 0)$. (05 Marks)
- c. Discuss types of data in cluster Analysis. (05 Marks)

- 6 a. What is cluster analysis? Explain the different types of clustering and clusters. (10 Marks)
b. Explain K – medoids clustering technique with algorithm. (10 Marks)
- 7 a. Brief the data mining applications. (10 Marks)
b. How do choose a data mining system? (10 Marks)
- 8 Write short notes on the following :
a. Schemas for multidimensional data. (08 Marks)
b. K – nearest neighbor classifier. (05 Marks)
c. Association rule mining. (07 Marks)