AUTUMN END SEMESTER EXAMINATION-2017

1st Semester M.Tech / Ph.D

DATA MINING & DATA WAREHOUSING

CS-6301

(Regular-2017 Admitted Batch)

Time: 3 Hours | Full Marks: 50

Answer any SIX questions including question No.1 which is compulsory.
The figures in the margin indicate full marks.
Candidates are required to give their answers in their own words as far as practicable and all parts of a question should be answered at one place only.

1. (a) What is difference between database and data warehouse? [1 × 10]

   (b) Define association and correlation.

   (c) Differentiate between data characterization and data discrimination.

   (d) Compare the advantages of FP growth algorithm over Apriori algorithm.

   (e) Compare drill down with roll up approach.

   (f) Justify that classification is a supervised learning.

   (g) What are the fields in which clustering techniques are used?

   (h) What is Regression? How it is useful?

   (i) Differentiate between metadata and data mart.

   (j) Distinguish STARjoin and STARindex.

2. (a) Demonstrate in detail about data mining steps in the process of knowledge discovery. [4]

(b) Consider the given set of data X ={15, 27, 62, 35, 39, 50, 44, 44, 22, 98}. Do preprocessing using Smoothing by bin means and bin boundary to smooth the data, using a bin of depth 3. Comment on the effect of this technique for the given data. [4]

3. (a) What are the basic statistical descriptions of the data? Briefly describe with assuming a suitable group of data. [4]

(b) What are the major steps involved in data preprocessing? Briefly describe them. [4]

4. (a) What are various OLAP operations? How it is different form OLTP? [4]

(b) Describe the design and construction of Data warehouses and explain with three-tier architecture diagram. [4]

5. (a) What is attribute selection measures? What are parameters taken in C4.5 algorithm to find the splitting attribute? Explain with a suitable example. [4]

(b) Explain Bayesian classification method with suitable example. [4]

6. (a) What are the objectives of cluster analysis? Explain the agglomerative hierarchical clustering algorithm with a suitable example. [4]

(b) What is grid based clustering? With an example explain an algorithm for grid based clustering. [4]

7. (a) A data base has following TEN transactions. [4]

Trans ID	Items Purchased
101	Milk, bread, eggs
102	Milk, Juice
103	Juice, butter
104	Milk, bread, eggs
105	Coffee, eggs
106	Coffee
107	Coffee, Juice
108	Milk, bread, eggs, cookies
109	Cookies, butter
110	Milk, bread

Find all frequent item sets using Apriori algorithm. Use 0.3 for the minimum support value. Illustrate each step of Apriori algorithm.

(b) Consider five points {X1, X2,X3, X4, X5} with the following coordinates as a two dimensional sample for clustering: X1 = (0,2.5); X2 = (0,0); X3= (1.5,0); X4 = (5,0); X5 = (5,2). Illustrate the k-means partitioning algorithm using the given data set. [4]

8. Short notes (Any TWO) [4 × 2]

(a) Outlier Analysis

(b) k-medoid Clustering

(c) Support Vector Machine

– ***** –