Total number of printed pages – 2

B. Tech
PEIT 5302

Fifth Semester Back Examination – 2014

DATA MINING AND DATA WAREHOUSING

BRANCH : IT

QUESTION CODE : L 301

Full Marks – 70

Time : 3 Hours

Answer Question No. 1 which is compulsory and any five from the rest.
The figures in the right-hand margin indicate marks.

1. Answer the following questions :   2 ×10
    (a) Differentiate between Fact and Dimension table.
    (b) Define Data Characterization and Data Discrimination.
    (c) Differentiate between Classification and Prediction.
    (d) What is Entity Identification Problem ?
    (e) Differentiate between Spatial mining and Temporal Mining.
    (f) What is an Outlier ? What do you mean by Outlier Mining ?
    (g) Define Binning. Name three binning strategies.
    (h) What is the difference between Symmetric and Asymmetric Binary variables ?
    (i) Differentiate between Supervised and Unsupervised learning.
    (j) What is a Decision tree ? How it is useful ?

2. (a) Define Data Mart and Data Warehouse. How is a data warehouse different from a database ? How they are similar ?   6
    (b) Describe the difference among the following approaches for the integration of a data mining system with a database or data warehouse system : No coupling, Loose coupling, Semitight coupling and Tight coupling.   4

3. (a) Use the two methods below to normalize the following group of data :   6
             200, 300, 400, 600, 1000 (v = 300, 600)
        (i) Min-max normalization by setting min = 0 and max = 1
        (ii) Z-score normalization.

P.T.O.

(b) Explain various methods for detecting redundancy in numerical and categorical data. 4

4. (a) Differentiate between OLAP and OLTP. Explain different OLAP operations in Multi Dimensional Data Model with suitable example. 6
   (b) Briefly explain Multitier Data Warehousing Architecture. 4

5. (a) Describe K means clustering with an example. 5
   (b) Explain the major clustering methods used in data mining. 5

6. (a) A database has five transactions. Let min_sup=60% and min_conf=80%

TID	ITEMS
T001	{A,B,C,D,E,F}
T002	{G,B,C,D,E,F}
T003	{A,H,D,E}
T004	{A,U,C,D,F}
T005	{C,B, B, D, I, E}

Find all frequent itemsets using Apriori. 7

(b) Given two objects represented by the tuples (22,1,42, 10) and (20,0,36,8):
Compute Euclidean distance between the two objects.
Compute the Manhattan distance between the two objects.
Compute the Minkowski Distance between the two objects using q=3. 3

7. (a) Describe three attribute selection measures for generating splitting criteria. 5
   (b) What is decision tree induction ? Explain major steps of Decision Tree Induction. 5

8. Write short notes on : 2.5×4
   (a) Association Rule Mining
   (b) Data Preprocessing
   (c) Histogram
   (d) PAM.