Registration No. : | | | | | | | | | | |

Total number of printed pages – 2

B. Tech

PEIT 5302

Fifth Semester Regular Examination – 2014

DATA MINING AND DATA WAREHOUSING

BRANCH(S) : BIOTECH, IT

QUESTION CODE : H 224

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) How is a data warehouse different from a database ? How are they similar ?

   (b) Define data discretization.

   (c) Why data transformation is essential in the process of knowledge discovery ?

   (d) Point out the major difference between the star schema and the snowflake schema ?

   (e) What are multidimensional database ? How do they play an important role in data warehouse ?

   (f) What is a decision tree ? What are the measuring factors of data mining ?

   (g) Define factless fact table with an example.

   (h) Differentiate between Spatial mining and Temporal mining.

   (i) Define support and confidence in Association rule mining.

   (j) Define nominal, ordinal and ratio scaled variables.

2. (a) What is data ware housing ? List out the characteristics of data warehouse. 5

   (b) Give the architecture of data warehouse and explain it. 5

3. (a) What are the types of data pre-processing techniques ? Explain, in detail, about them. 5

P.T.O.

(b) Suppose that the association rule *"hot dogs=>hamburgers"* is mined. Given a minimum support threshold of 25% and a minimum confidence threshold of 50%, is this association rule strong ? What kind of *correlation* relationship exists between the two ? 5

	HOTDOGS	$\overline{\text{HOTDOGS}}$	$?_{ROW}$
HAMBURGERS	2,000	500	2,500
$\overline{\text{HAMBURGERS}}$	1,000	1,500	2,500
$?_{COL}$	3,000	2,000	5,000

4. (a) Differentiate OLAP and OLTP. Explain with suitable example. Give an account of OLAP tools. 5

(b) What is the role of Data Mart in the administration and management of data warehouse ? Explain. 5

5. (a) What is decision tree ? Explain how classification is done using decision tree induction. 5

(b) Write and explain the algorithm for mining frequent itemsets without candidate generation. 5

6. (a) Why is outlier mining important ? Briefly describe the different approaches behind deviation based outlier detection. 5

(b) Describe the various descriptive statistical measures for data mining. 5

7. Suppose that the data mining task is to cluster the following eight points (with (x, y) representing location) into three clusters : 10

$A1(2, 10)$, $A2(2, 5)$, $A3(8, 4)$, $B1(5, 8)$, $B2(7, 5)$, $B3(6, 4)$, $C1(1, 2)$, $C2(4, 9)$:

The distance function is Euclidean distance. Suppose initially we assign $A1$, $B1$ and $C1$ as the center of each cluster, respectively. Use the k-means algorithm to show only (a) The three cluster centers after the first round execution, (b) The final three clusters.

8. Write short notes on any two : 5×2

(a) BIRCH

(b) CART

(c) FACT CONSTELLATION SCHEMA

(d) ROLAP.