AUTUMN END SEMESTER EXAMINATION-2018

7th Semester B.Tech

DATA MINING AND DATA WAREHOUSING

IT 4037

[For 2016(L.E.) & 2015 Admitted Batches]

Time: 3 Hours | Full Marks: 60

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 market basket analysis? [2 × 10]

   (b) Define Gain ratio and Entropy in a classifier.

   (c) Write down the difference between OLAP and ROLAP.

   (d) Define ETL process of data cleaning.

   (e) Define support and confidence.

   (f) Consider the one dimensional dataset shown below:

x	0.5	3.0	4.5	4.6	4.9	5.2	5.3	5.5	7.0	9.5
y	-	-	+	+	+	-	-	+	-	-

   Classify the data point x=5.0 according to its 3-Nearest Neighbours (using majority vote).

   (g) What is data cube? Define slice and dice operation.

   (h) What is an outlier analysis?

   (i) What is Lazy learner? Why it is used?

   (j) How Data warehouse is different from operational database?

2. (a) Discuss the Data mining task with suitable examples. [4]

(b) Briefly explain any five data pre-processing approaches. [4]

3. (a) What is decision tree? Why pruning is required? Explain with an example. [4]

(b) Discuss the star and snowflake schema in detail with suitable example. [4]

4. (a) What is knowledge discovery in database (KDD)? [4]

(b) Describe the k-means and k-mediods algorithm in terms of shape of cluster that can be determined and parameter that must be specified? [4]

5. (a) With a neat diagram explain the components of data warehouse. [4]

(b) Describe the different methods for data cleaning. [4]

6. (a) Explain the distinction between [2×2]

(i) Measures and Dimensions

(ii) Fact tables and Dimension tables.

(b) What is the difference between Agglomerative and divisive clustering? Explain. [4]

7. A database has nine transactions let minimum support = 30% and find all frequent item sets using the above algorithm also find the association rule using Apriory algorithm. [8]

TID	LIST OF ITEMS
1	A,B,E
2	B,D
3	B,C
4	A,B,D
5	A,C
6	B,C
7	A,C
8	A,B,C,E
9	A,B,C

8. Write short notes on *any two*. [4×2]

(a) Temporal mining

(b) Spatial mining

(c) Neural network classifier

(d) Genetic Algorithm

*****