PVP 14

Code: IT6T4

III B.Tech - II Semester – Regular Examinations – May 2017

DATA MINING & DATA WAREHOUSING
(INFORMATION TECHNOLOGY)

Duration: 3 hours Max. Marks: 70

PART – A

Answer *all* the questions. All questions carry equal marks

11x 2 = 22 M

1.

a) Define the term 'Data Mining'.
b) What can you say about normalization and list out its types?
c) What is 'OLAP'?
d) Can you list various operations of OLAP?
e) What approach would you use to mine association rules from large databases?
f) How would you explain closed frequent item set and maximal frequent item set?
g) What is the main idea of naïve Bayesian classification?
h) How would you compare eager classification and lazy classification?
i) Can you list major clustering methods?
j) Can you make a distinction between symmetric and asymmetric binary variables?
k) What are the characteristics of data warehouse?

PART – B

Answer any *THREE* questions. All questions carry equal marks.

3 x 16 = 48 M

2. a) Can you explain the functionalities of data mining? 12 M

   b) How would you justify that all of the patterns are interesting? 4 M

3. a) Explain in detail the process of designing a data ware house. 8 M

   b) Explain the components in data warehouse architecture models. 8 M

4. How could you determine frequent item sets with their support for the following transactional database? (Let the minimum support be 50% and confidence be 60%). 16 M

TransID	**Items**
T100	A, B, C, D
T200	A, B, C, E
T300	A, B, E, F, H
T400	A, C, H

5. Can you predict the outcome decision tree for the following test data? Risk is the class attribute. Generate all IF-THEN rules from the final decision tree. 16 M

Gender	Height	Risk
F	1.5	LOW
M	1.9	HIGH
F	1.8	MEDIUM
F	1.8	MEDIUM
F	1.6	LOW
M	1.8	MEDIUM
F	1.5	LOW
M	1.6	LOW
M	2.0	HIGH
M	2.0	HIGH

6. a) How would you explain the following clustering methods
i) OPTICS ii) DENCLUE 8 M

b) Can you make a distinction between K-means and DBSCAN clustering methods. 8 M