USN | | | | | | | | | | |

10BBI23

Second Semester M.Tech. Degree Examination, June 2012
Data Warehousing and Data Mining

Time: 3 hrs. Max. Marks:100

Note: ***Answer any FIVE full questions.***

1 a. What is a data warehouse? Why data warehouse is required? What are the characteristics of data warehouse? **(10 Marks)**

b. Explain the three – tier architecture of a data warehouse. **(10 Marks)**

2 a. List the various methods of data modeling and explain the star schema, with an example. **(10 Marks)**

b. Distinguish between data ware house and operational database system. **(10 Marks)**

3 a. What is OLAP? Explain its operations. **(10 Marks)**

b. Discuss the different types of OLAP servers. **(10 Marks)**

4 a. What is data mining? Why data preprocessing is required? Explain how inconsistency is removed. **(10 Marks)**

b. Describe all the steps of KDD process. **(10 Marks)**

5 a. What is CRISP – DM? Why it is essential? What are the benefits derived from CRISP. **(10 Marks)**

b. What are inputs, instances and concept? Explain the classification of attributes with examples. **(10 Marks)**

6 a. What is classification? Explain the purposes of classification model. Also explain the Hunt's algorithm. **(10 Marks)**

b. For the following data sets construct a decision tree split, based on outlook and compute the entropy to determine the best split. **(10 Marks)**

Item Number	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Rainy	Mild	High	False	Yes
5	Rainy	Cool	Normal	False	Yes
6	Rainy	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Sunny	Mild	High	False	No
9	Sunny	Cool	Normal	False	Yes
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Rainy	Mild	High	True	No

7 a. What is clustering? Explain the major clustering methods. **(10 Marks)**

b. What is Apriori algorithm? A database has four transactions. Let Min_ Sup = 40% and min_count = 60%.

TID	Date	Items bought
100	10/15/11	{K, A, B, D}
200	10/15/11	{D, A, C, E, B}
300	10/19/11	{C, A, B, E}
400	10/22/11	{B, A, D}

Find all frequent items using Apriori algorithm. **(10 Marks)**

8 Write short notes on :

a. Regression.

b. DMQL.

c. Cross – validation.

d. Error measures. **(20 Marks)**

*****