Total Number of Pages: 2   <u>**B.TECH**</u>
                                                               PEIT5302

5th Semester Regular / Back Examination 2015-16
DATA MINING AND DATA WAREHOUSING
BRANCH: BIOTECH,IT,TEXTILE
Time: 3 Hours
Max Marks: 70
Q.CODE: T674
Answer Question No.1 which is compulsory and any five from the rest.
The figures in the right hand margin indicate marks.

Q1   Answer the following questions: **(2 x 10)**
   a) Differentiate between spatial mining and temporal mining.
   b) Differentiate between classification and clustering.
   c) Define Support and Confidence in reference to Association Rule mining.
   d) What is strategic information?
   e) What is data granularity?
   f) Why we need a separate data staging area in a data warehouse?
   g) Compare two-tier and three-tier architecture of a data warehouse.
   h) Name the different soft ware tools used in a data ware house.
   i) What is dimensional modeling?
   j) Explain why ETL functions are very challenging?

Q2 a) Explain why operational systems are not useful for making strategic decisions?   **(5)**
   b) Why is data cleansing and data transformation functions considered to be a vital task in the integration process   **(5)**

Q3 a) Explain how a data warehouse differs from a data mart?   **(5)**
   b) Compare the top-down and bottom-up approach of designing data marts.   **(5)**

Q4 a) Explain the key issues to be considered while planning for a data warehouse.   **(5)**
   b) List down the features of a fact table and a dimension table.   **(5)**

Q5 a) Give a comparative analysis between Star schema, Snow flake schema and fact constellations.   **(5)**
   b) Why is metadata used for data acquisition, data storage and information delivery?   **(5)**

Q6 **a)** Explain the steps involved in KDD process? How is data mining **(5)** different from KDD?

b) Explain the major clustering methods used in data mining. **(5)**

Q7 **a)** Describe the different types of OLAP tools. **(5)**

b) Explain the Apriori algorithm for mining frequent itemsets **(5)**

Q8 Write short notes on any two: **(5 x 2)**

a) Outlier analysis

b) Decision tree induction

c) Partition around Mediod

d) Issues in data mining