

2014  
**B.E. (Computer Science and Engineering)**  
**Sixth Semester**  
**Elective – I**  
**CS-605C: Data Mining and Analysis**

Time allowed: 3 Hours

Max. Marks: 50

**NOTE: Attempt five questions in all, including Question No. 1 which is compulsory and selecting two questions from each Part.**

x-x-x

- I** a) Mention any two issues related to user interaction in a data mining system.  
 b) Distinguish between classification and clustering.  
 c) Give some examples of the kind of multidimensional analyses that can be done on spatiotemporal databases.  
 d) What are the advantages and limitations of using a MOLAP server?  
 e) Discuss any 2 measures of data quality. (5X2)

**Part A**

**II** Suppose a data mart has to be designed for engineering department of a University. It should store data about students, faculty, courses, infrastructure. Analyses of student placements, and academic performance needs to be done frequently.

- a) Draw the appropriate multidimensional schema diagram for this data mart. (7)  
 b) Starting with the base cuboid, what OLAP operations should be performed to list the average package offered to CSE students in the year 2022. (3)

- III** a) Discuss any two techniques for data transformation. (5)  
 b) Differentiate between characterization and discrimination with the help of suitable example. (5)

**IV** Distinguish between RDBMS, data mart and data warehouse. (10)

**Part B**

**V** Describe how naïve bayes theorem can be used for classification. State the advantages and limitations of this technique. (10)

**VI** Explain how mining of single level associations with candidate generation can be done from transaction database. (10)

**VII** Write short notes on:

- a) OLAM  
 b) Mining text databases (5, 5)

x-x-x