**Roll No.** ⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚                **Total No. of Pages : 02**

**Total No. of Questions : 08**

**M.Tech.(IT) E2 (2015 & Onwards)/(CSE Engg.) EI-II (2015 to 2017)**
**(Sem.–2)**
# DATA WAREHOUSING AND DATA MINING
## Subject Code : MTCS-207
## Paper ID : [72891]

**Time : 3 Hrs.**                                    **Max. Marks : 100**

**INSTRUCTION TO CANDIDATES :**
 1.  **Attempt any FIVE questions out of EIGHT questions.**
 2.  **Each question carry TWENTY marks.**

1.  There are several typical cube computation methods, such as Multi-Way, BUC, and Star-Cubing. Describe any two of these methods and compare their feasibility and performance under the following conditions :

    a) Computing a dense full cube of low dimensionality (e.g., less than eight dimensions).

    b) Computing an iceberg cube of around 10 dimensions with a highly skewed data distribution.

2.  a) Briefly describe the following approaches to clustering  partitional  methods  and density-based methods. Give examples in each case.

    b) What is a Confusion Matrix for classifier? Write a 4×4 confusion matrix for a classifier with 100% accuracy. The instances in the four classes are A(20), B(35), C(40), D(10). Also calculate classifier precision in detail?

3.  There are several major differences between biological sequential patterns and transactional sequential patterns. First, in transactional sequential patterns, the gaps between two events are usually nonessential. For example, the pattern "*purchasing a digital camera two months after purchasing a PC*" does not imply that the two purchases are consecutive. However, for biological sequences, gaps play an important role in patterns. Second, patterns in a transactional sequence are usually precise. However, a biological pattern can be quite imprecise, allowing insertions, deletions, and mutations. Discuss how the mining methodologies in these two domains are influenced by such differences.

4.  a) Describe the techniques for improving the efficiency of DHP-based mining.

    b) How is the performance of various mining algorithms evaluated? Explain.

5. a) What are the different methods used for the estimation of predictive accuracy of various classification methods?

   b) How is multi-dimensional analysis in multimedia data conducted? Explain.

6. a) From the architecture point of view, there are three data warehouse models : enterprise warehouse, data mart and virtual warehouse. Explain Data mart and Virtual warehouse,

   b) What are Bayesian classifiers? Explain briefly Baye's theorem. Also explain how Naive Bayesian classifier works.

7. Explain the following :

   a) Search engine architecture.

   b) Enterprise search.

8. a) Simple classification of web mining is '*Web Content Mining*' and '*Web Usage Mining*'. Write a short note on '*Web Content Mining*'.

   b) Explain FP growth algorithm for mining Association Rule in large databases.