Code No: **R41053**      **R10**      **Set No. 1**

### IV B.Tech I Semester Supplementary Examinations, March - 2017
## DATA WAREHOUSING AND DATA MINING
### (Common to Computer Science & Engineering and Information Technology)

**Time: 3 hours**      **Max. Marks: 75**

**Answer any FIVE Questions**
**All Questions carry equal marks**
**\*\*\*\*\***

1 a) What is the need for preprocessing? List various preprocessing techniques? [8]

  b) Explain the steps in the data mining process giving example for each step? [7]

2 a) Explain similarity between binary vectors? [6]

  b) What is Jaccard coefficient? Derive an expression for Jaccard coefficient? [9]

3 a) Explain the star schema of data ware house model with the help of a diagram? [6]

  b) Differentiate between data ware house and operational database? [7]

4 a) Explain the decision tree classifier with the help of a diagram? [8]

  b) Explain over fitting and tree pruning? [7]

5 a) Explain the nearest neighbor algorithm for classification? [8]

  b) Explain the parameters for the evaluation of classifier? [7]

6 a) Explain any association mining algorithm without generating candidate item sets? [9]

  b) Define support and confidence? What is support threshold? [6]

7 a) What are the advantages of k-means algorithm? [8]

  b) What are the additional issues of k-means algorithm? [7]

8 a) Explain the basic hierarchical clustering strategies, AGNES and DIANA? [7]

  b) Explain DB SCAN clustering algorithm in detail? [8]