Subject Code: MC1343/R13

M C A - IV Semester Regular Examinations, June/July - 2015
DATA WAREHOUSING AND MINING

Time: 3 hours   Max Marks: 60

Answer any FIVE of the following
All questions carry equal marks
****

1. What is Data Mining? On what type of data, data mining can be performed? With an example, briefly describe any two similarity/dissimilarity techniues for continuos data.

2. a). What is classification? For an example data, briefly describe the process of decision tree induction.
   b). Briefly describe the methods for expressing attribute test conditions. Why these tests require?

3. What is model over fitting? Briefly describe the reasons for model over fitting? List the available solutions for solving the over fitting.

4. What is an impurity measure? With examples, briefly describe any three measures for selecting the best split.

5. a). What is FP Growth? With an example, briefly describe the process of constructing an FP-tree.
   b). What are maximal frequent itemsets? Brielfy describe closed frequent itemsets.

6. What is clustering? What is DBSCAN? What are strength and weaknesses of DBSCAN? Briefly describe DBSCAN technique.

7. What is proximity matrix? Briefly describe any one supervised measure for obtaining cluster validity.

8. a). What is a search engine? Briefly describe the characterstics of search engines.
   b). Briefly describe the process of ranking webpages.

****

`