AUTUMN END SEMESTER EXAMINATION-2018

1st Semester M.Tech

DATA MINING AND DATA WAREHOUSING

CS 6301

[For 2018 & 2017 Admitted Batches]

Time: 3 Hours | Full Marks: 50

Answer any SIX questions including question No.1 which is compulsory.
The figures in the margin indicate full marks.
Candidates are required to give their answers in their own words as far as practicable and all parts of a question should be answered at one place only.

1. (a) Differentiate between OLAP and OLTP. [1 × 10]

   (b) What are the different types of metadata?

   (c) When we can say the association rules are interesting?

   (d) What are the rule strength measures used in any association rule model? Discuss with their probability formula.

   (e) Why hyperplane is used in SVM?

   (f) Consider the following set of frequent 3-itemsets: {1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 3, 5}, {2, 3, 4}, {2, 3, 5}, {3, 4, 5}. Assume that there are only five items in the data set. List all candidate 4-itemsets obtained by a candidate generation procedure.

   (g) Why "Bias" is used in ANN?

   (h) What are the different issues in clustering?

   (i) Explain the definition of a centroid in k-means.

   (j) What is Time Series Analysis?

2. (a) Explain three tier Data Warehousing Architecture. [4]

(b) Explain different types of schemas used in multidimensional model in data warehousing. [4]

3. (a) Suppose the confidence of the rules A → B and B → C are larger than some threshold, *minconf (50%)*. Is it possible that A → C has a confidence less than *minconf*? Justify. [4]

[Hints: s(A,B) = 60%, s(A) = 90%, s(A,C) = 20%, s(B) = 70%, s(B,C) = 50%, s(C) = 60%]

(b) Describe the different dimensions that can be used to assess the quality of data. [4]

4. (a) State Bayes theorem and discuss how Bayesian classifier works. [4]

(b) Information gain act as an attribute selection measure in decision tree, justify. [4]

5. Consider the following database of houses represented by 5 training examples. The target attribute is 'Acceptable', which can have values 'yes' or 'no'. This is to be predicted based on the other attributes of the house. [8]

House	Furniture	Nos. rooms	New kitchen	Acceptable
1	No	3	Yes	Yes
2	Yes	3	No	No
3	No	4	No	Yes
4	No	3	No	No
5	Yes	4	No	Yes

i. Construct the decision tree from the above table, that would be learned by the ID3 algorithm.

ii. Show the value of the information gain for each candidate attribute at each step in the construction of the tree.

6. (a) Discuss various methods of cluster analysis with suitable example for each. [4]

(b) What is Business Intelligence and How it is related to DataMining? [4]

7. (a) Consider five points {x1, x2, x3, x4, x5} with the coordinates as a two dimensional sample for clustering: x1 = (0,2.5); x2 = (0,0); x3= (1.5,0); x4 = (5,0); x5 = (5,2). Illustrate the k-means partitioning algorithm for next iteration using the given data setwith k=2 and initially, clusters are formed from random distribution of samples: C1 = {x1, x2, x4} and C2 = {x3, x5}. [4]

(b) Discuss DBSCAN clustering algorithm with a suitable example. [4]

8. Answer any **TWO**. [4×2]

(a) KDD

(b) FP Tree Construction

(c) Hierarchical clustering

*****