CBCS SCHEME

Sixth Semester B.E. Degree Examination, Dec.2019/Jan.2020
Data Mining and Data Warehousing

Time: 3 hrs.   Max. Marks: 80

Note: *Answer any FIVE full questions, choosing ONE full question from each module.*

Module-1

1 a. What is Data warehouse? Explain three tier architecture of data warehouse. (08 Marks)
   b. Explain the schemas of multidimensional data models. (08 Marks)

OR

2 a. What is Data cube measure? Explain the categorization of measures. (08 Marks)
   b. Explain data cube operations with examples. (08 Marks)

Module-2

3 a. Explain data cube computation and curse of dimensionality. (08 Marks)
   b. Explain different methods of indexing OLAP data. (08 Marks)

OR

4 a. State and explain various data mining tasks. (08 Marks)
   b. Define Similarity and dissimilarity between the objects. Find SMC and Jaccord's coefficient of two binary vectors.
   X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)   Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1). (08 Marks)

Module-3

5 a. What is Association Analysis? Explain Association rule , Support and Confidence. (08 Marks)
   b. State Apriori principle. Write apriori algorithm for frequent itemset. (08 Marks)

OR

6 a. Construct an FP tree for the following dataset.

TID	Items
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}
4	{a, d, e}
5	{a, b, c}
6	{a, b, c, d}
7	{a}
8	{a, b, c}
9	{a, b, d}
10	{b, c, e}

(08 Marks)

   b. Explain the strategies used in frequent itemset generation. (08 Marks)

Module-4

7 a. Explain the general approach for solving classification problem. (08 Marks)
   b. Write the algorithm for decision tree induction. (08 Marks)

OR

8 a. Explain the methods of comparing classifiers. (08 Marks)
   b. Write the characteristics of nearest neighbor classifier. (08 Marks)

Module-5

9 a. Explain the requirements of cluster analysis. (08 Marks)
   b. State and explain K – means algorithm. (08 Marks)

OR

10 a. Write DBSCAN clustering algorithm and estimate time and space complexity. (08 Marks)
    b. State and explain the issues in cluster evaluation. (08 Marks)

* * * * *