CBCS SCHEME

USN | | | | | | | | | | |

16/17MCA442

Fourth Semester MCA Degree Examination, Jan./Feb. 2021
Data Warehousing and Data Mining

Time: 3 hrs. Max. Marks: 80

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

Module-1

1 a. Define a Datawarehouse. Give 3 imporatant benefits of implementing a datawarehouse for a large enterprise. **(08 Marks)**

b. Explain the most imporatant guidelines in implemeing a datawarehouse. **(08 Marks)**

OR

2 a. Explain FASMI characterstics of OLAP systems. **(04 Marks)**

b. Compare between OLTP and OLAP. **(04 Marks)**

c. What is ETL? Explin ODS and DW architecture. **(08 Marks)**

Module-2

3 a. What is data mining? With a neat diagram explain the KDD process. **(08 Marks)**

b. Define Proximity for the following vectors X and Y, calculate the similarity measure :
X = (1, 1, 0, 1, 0, 1) ; Y = (1, 1, 1, 0, 0, 1)
i) Cosine ii) Jaccard iii) SMC. **(08 Marks)**

OR

4 a. List and explain the different types of attributes with example. **(08 Marks)**

b. Explain the specific challenges that motivated the development of data minig. **(08 Marks)**

Module-3

5 a. State FP-growth algoirthm. Construct FP-tree for the following transaction data set.

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. Define Apriori principle. Brielfy discuss Arrrori algoirhtm for frequent itemset generation. **(08 Marks)**

OR

6 a. Define Maximal frequent itemset and closed frequent itemsets. **(08 Marks)**
b. What is an association rule? Define : i) Support ii) Confidence iii) Frequent itemset. **(08 Marks)**

Module-4

7 a. Explain Rule based classifiers with examples. **(08 Marks)**
b. Discuss on the estimation of predictive accuracy of classification methods. **(08 Marks)**

OR

8 a. What is the impurity for the node N having one element in class 1 and 5 elements in class 2 as measured by entropy, Gini index and classification error? **(08 Marks)**
b. Explain K – NN classification algorihtm with characterstics of NN- classifiers. **(08 Marks)**

Module-5

9 a. List and explain the desired features of cluster analysis. **(08 Marks)**
b. What is cluster Analysis? Explain the basic K-means clustering technique. **(08 Marks)**

OR

10 a. Explain how the quality of a clustering algorithm is determined. **(08 Marks)**
b. How can the distance between a pair of points be computed? List the features of each distance metric. **(08 Marks)**

* * * * *