CBCS SCHEME

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

16/17MCA442

Fourth Semester MCA Degree Examination, Aug./Sept.2020
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. Describe the data warehouse architecture. **(08 Marks)**

b. What is data warehouse? Explain various data warehouse modeling? **(08 Marks)**

OR

2 a. List and explain any 4 characteristics of OLAP? **(08 Marks)**

b. Explain OLAP operations in detail with an example. **(08 Marks)**

Module-2

3 a. Brief the major tasks in data pre-processing. **(05 Marks)**

b. Explain Knowledge Discovery (KDD) process with neat diagram. **(06 Marks)**

c. Define an Attribute. Explain the different types of Attributes. **(05 Marks)**

OR

4 a. List and explain any 4 challenges of data mining. **(08 Marks)**

b. Brief the important characteristics of structured data. **(08 Marks)**

Module-3

5 a. What is association Rule Mining? Write an APRIORI algorithm for finding a frequent item set. **(08 Marks)**

b. Explain Compact representation of frequent item set. **(06 Marks)**

c. Define association Rules. **(02 Marks)**

OR

6 a. Describe the alternative methods for frequent item set generation. **(08 Marks)**

b. Construct the FP-tree for the following 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	{B, C}
8	{A, B, C}
9	{A, B, D}
10	{B, C, E}

(08 Marks)

Module-4

7 a. Explain different methods for expressing Attribute test condition. **(08 Marks)**
b. Discuss on measure for selecting best split. **(06 Marks)**
c. Define confusion matrix. **(02 Marks)**

OR

8 a. With an example, explain multiclass problem. **(06 Marks)**
b. Explain the different methods used for estimating predictive accuracy of classification. **(10 Marks)**

Module-5

9 a. Discuss the different type of data in clustering anlaysis. **(08 Marks)**
b. Briefly discuss the commonly used partitioning methods. **(08 Marks)**

OR

10 a. Briefly discuss the different clustering approaches. **(10 Marks)**
b. Write a Density based clustering or DB Scan algorithm. **(06 Marks)**

* * * * *