CBCS SCHEME

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

16MCA442

Fourth Semester MCA Degree Examination, Dec.2018/Jan.2019

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. What is a data warehouse? With a neat diagram, explain three-tier data warehousing architecture. **(08 Marks)**

b. With a neat diagram explain star schema and snowflake schema diagram for the data warehouse. **(08 Marks)**

OR

2 a. Discuss the different data warehouse models and also explain the recommended approach for data warehouse development. **(08 Marks)**

b. Discuss about the typical OLAP (online analytical processing) operations on multidimensional data with an example. **(08 Marks)**

Module-2

3 a. What is data mining? Explain the process of knowledge discovery in databases (KDD) with a neat diagram. **(10 Marks)**

b. Explain the different strategies for data transformation. **(06 Marks)**

OR

4 a. Describe the different methods for data cleaning. **(10 Marks)**

b. Explain the following terms:
i) Principal components analysis
ii) Histograms **(06 Marks)**

Module-3

5 a. Define minimum support and minimum confidence. Write the Aprori algorithm for frequent item set mining. **(09 Marks)**

b. Derive the frequent item set for the below transactional data given the minimum support as 50% and minimum confidence as 60%. **(07 Marks)**

TID	Items
T_{101}	I_1, I_3, I_4
T_{102}	I_2, I_3, I_5
T_{103}	I_1, I_2, I_3, I_5
T_{104}	I_2, I_5

OR

6 a. Explain the various measures of evaluating association patterns. **(09 Marks)**

b. Write FP-growth algorithm for discovering frequent item sets without candidate generation. **(07 Marks)**

Module-4

7 a. Explain the steps involved in Naive's Bayesian classification. **(10 Marks)**

b. How rule based classifiers are used for classification? Explain. **(06 Marks)**

OR

8 a. Explain classification process. Write an algorithm for decision tree induction technique. **(09 Marks)**

b. List the different characteristics of decision tree induction. **(03 Marks)**

c. Write short note on:

i) Bagging

ii) Boosting **(04 Marks)**

Module-5

9 a. Explain desired features of cluster analysis. **(09 Marks)**

b. Explain K-means clustering method and algorithm. **(07 Marks)**

OR

10 a. Write short notes on:

i) BIRCH

ii) Chameleon **(08 Marks)**

b. Describe the DBSCAN algorithm for clustering. **(08 Marks)**

* * * * *