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

13MCA442

Fourth Semester MCA Degree Examination, Dec.2018/Jan.2019
Data Warehousing and Data Mining

Time: 3 hrs. Max. Marks:100

Note: *Answer any FIVE full questions.*

1 a. Define data warehouse. What are its key features? Explain and compare between OLTP and OLAP systems. **(10 Marks)**

b. Explain multidimensional data model for data warehouse and also schemas. **(10 Marks)**

2 a. What is data mining? Explain motivating challenges for the development of datamining. **(10 Marks)**

b. Explain the applications of datamining in various fields. **(10 Marks)**

3 a. Explain different types of datasets with example. **(10 Marks)**

b. What are the techniques used for feature subset selection? Explain the architecture for feature subset selection. **(10 Marks)**

4 a. Write Apriori algorithm for frequent itemset generation and mention its principle. Generate frequent itemset using Apriori algorithm for the following market basket transaction.

TID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}

Assume support threshold is 60% i.e. minimum support count is equal to 3. **(10 Marks)**

b. Construct FP tree for the transaction data set shown in table Q4 (b) and explain the steps of construction using FP Growth Algorithm. **(10 Marks)**

Transaction 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}

Table Q4 (b)

5 a. Write and explain decision tree induction along with Hunts Algorithm. **(10 Marks)**

b. Explain sequential covering algorithm and rule-growing strategy for rule extraction. **(10 Marks)**

6 a. Write algorithm for K-nearest-neighbor classification and explain nearest neighbor classifiers. **(10 Marks)**

b. Explain multiclass problem and its approaches. **(10 Marks)**

7 a. What is cluster analysis? Explain different types of cluster analysis methods in brief. **(10 Marks)**

b. What do you mean by density based clustering? Write algorithm for DBSCAN and explain the classification points according to centre based density. **(10 Marks)**

8 a. Explain statistical approaches for outlier analysis and its strengths and weakness. **(10 Marks)**

b. Explain clustering based techniques for outlier analysis and its strengths and weakness. **(10 Marks)**

* * * * *

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.