CBCS Scheme

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



16MCA442

# Fourth Semester MCA Degree Examination, June/July 2018
## 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. Explain the three types of schemas in multidimensional data model with example. **(08 Marks)**

b. Describe 3 – tier datawarehouse architecture with a neat diagram. **(08 Marks)**

**OR**

2 a. Explain data cube operations with example for each operation. **(08 Marks)**

b. What is Datawarehouse? Compare OLTP and OLAP systems. **(08 Marks)**

### Module-2

3 a. What is Data Mining? Explain KDD process in data mining with a neat diagram. **(08 Marks)**

b. Briefly explain motivating chllenges in the field of data mining. **(04 Marks)**

c. Briefly discuss various applications of data mining. **(04 Marks)**

**OR**

4 a. For the below given 2×2 contingency table with two attributes "gender" and "preferred reading" conduct correlation analysis between the given attributes using Chi-square test.
Note: Expected frequencies are given inside parentheses.

| | | Gender | | |
|---|---|---|---|---|
| | | Male | Female | Toal |
| Referred - reading | Fiction | 250 (90) | 200(360) | 450 |
| | Non fiction | 50(210) | 1000(840) | 1050 |
| | Total | 300 | 1200 | 1500 |

**(08 Marks)**

b. Briefly explain Min – Max, Z – score and Normalization by decimal scaling with a suitable example. **(08 Marks)**

### Module-3

5 a. Define Apriori principle, briefly discuss. Apriori algorithm for Frequent Itemset generation. **(08 Marks)**

b. For a given transaction data, generate frequent itemset and identify valid Association Rules with minimum support as 60% and minimum confidence as 75%

| $T_{id}$ | Items |
|---|---|
| 1 | Bread, Cheese, eggs, Juice |
| 2 | Bread, Cheese, Juice |
| 3 | Bread, Milk, Yogurt |
| 4 | Bread, Juice, Milk |
| 5 | Cheese, Juice, Milk |

**(08 Marks)**

**OR**

6 a. Explain Maximal, frequency Intemset and closed Frequent Intemset techniques for compact representation. Using an example for each. **(08 Marks)**

b. Explain and construct FP Tree for given Transaction data :
Data set :

| $T_{id}$ | 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)**

## Module-4

7 a. Write and explain Hunt's algorithm with a suitable example. **(08 Marks)**

b. Explain rule based classifier technique with an example. **(08 Marks)**

**OR**

8 a. Write and explain K- nearest neighbor classification algorithm. **(08 Marks)**

b. Write a note on Naïve Bayer Classifiers. **(04 Marks)**

c. List out and explain any four Evaluation criteria for classification methods. **(04 Marks)**

## Module-5

9 a. Mention and explain the desired features of cluster analysis. **(06 Marks)**

b. Write a note on :
i) Manhattan distance
ii) Euclidean distance. **(06 Marks)**

c. Briefly explain different types of data used for data mining. **(04 Marks)**

**OR**

10 a. What is cluster analysis? Briefly explain different types of cluster analysis methods. **(08 Marks)**

b. Briefly explain DBSCAN algorithm in density based clustering. **(08 Marks)**

* * * * *