Following Paper ID and Roll No. to be filled in your Answer Book.

PAPER ID : 21211 | Roll No. | | | | | | | | |

MCA Examination 2018-19

(Even Semester)

DATAWARE HOUSING AND DATA MINING

Time : Three Hours]   *[Maximum Marks : 60*

Note :– Attempt all questions.

SECTION – A

1. Attempt all parts of the following :   8×1=8

   (a) What do you mean by dataware house?

   (b) What are different data pre-processing technique?

   (c) What is frequent item set?

   (d) Define association rule.

   (e) Define classification.

[P. T. O.

(f)   Give uses of clustering.

(g)   What are the use of data mining in finance sector?

(h)   Give few examples of data mining in retail industry.

SECTION – B

2.   Attempt any two parts of the following : $2 \times 6 = 12$

   (a)   Use the two methods below to normalize the following group of data :

$$200, \ 300, \ 400, \ 600, \ 1000$$

     (i)   Min - Max normalization by setting $min = 0$ and $max = 1$.

     (ii)   Z-score normalization.

   (b)   Transactional data for a store is as follows :

TID	List of items IDs
T_{100}	I_1, I_2, I_5
T_{200}	I_2, I_4
T_{300}	I_2, I_3
T_{400}	I_1, I_2, I_4

T_{500}	I_1, I_3
T_{600}	I_2, I_3
T_{700}	I_1, I_3
T_{800}	I_1, I_2, I_3, I_5
T_{900}	I_1, I_2, I_3

Using FP-grow algorithm find frequent item set from above transaction data if minimum support count = 2.

(TID → Set of transaction, I_N → item)

(c) Describe K-mean clustering with an example.

(d) How data mining provides help in biological data analysis?

SECTION–C

Note :– Attempt all questions from this section. 10×4=40

3. Attempt any two parts of the following :

(a) Describe different multidimensional data base schema with diagram.

(b) Describe different technique of data pre-processing.

[P. T. O.

(c) With suitable diagram describe thre-tier architecture of data warehouse.

4. Attempt any two parts of the following :

    (a) Describe nearest neighbour classifier algorithm.

    (b) How frequent item set are generated using vertical data formate? Explain with example.

    (c) What is Market Basket Analysis? Why it has been performed?

5. Attempt any two parts of the following :

    (a) How best set of clusters can be find? Explain.

    (b) Describe agglomerative hierarchical clustering.

    (c) How distance between clusters recorded?

6. Attempt any two parts of the following :

    (a) How data mining is useful for intrusion detection?

    (b) What are the social impact of data mining?

    (c) What are the guidelines for fair information practices?

※※※