Total number of printed pages–4

53 (IT 701) DMDW

2019

DATA MINING AND DATA WAREHOUSING

Paper : IT 701

Full Marks : 100

Time : Three hours

The figures in the margin indicate full marks for the questions.

*Answer **any five** questions.*

1. *(a)* Define Data Mining. Explain the data mining system architecture. 10

   (b) Draw a simple reference model for Knowledge Discovery in Databases (KDD) process and explain each entity briefly. 10

2. *(a)* Discuss the Apriori algorithm. 8

Contd.

(b) For the following given transaction data-set, generate rules using Apriori algorithm. Consider the values as Support = 50% and Confidence = 75%.

Transaction ID	Items Purchased
1	Bread, Cheese, Egg, Juice
2	Bread, Cheese, Juice
3	Bread, Milk, Yogurt
4	Bread, Juice, Milk
5	Cheese, Juice, Milk

12

3. Explain the KNN classification and perform the classification algorithm on the following data set and predict the class for $X(P_1 = 3, P_2 = 7)$, $K = 3$. 20

P_1	P_2	Class
7	7	False
7	4	False
3	4	True
1	4	True

4. Define cluster analysis. List and explain the applications of cluster analysis. Also explain various types of clustering. 20

5. Discuss *k*-means clustering algorithm. Using *k*-means clustering algorithm, divide the given dataset into two clusters. 20

Id	*X*	*Y*
1	1	1
2	1·5	2
3	3	4
4	5	7
5	3·5	5
6	4·5	5
7	3·5	4·5

6. For the following weather attributes, create decision tree. Sample data '*S*' is given below : 20

Day	Outlook	Temperature	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

7. Write short notes on the following : ***(any four)*** 5×4=20

(i) Datasets

(ii) Data Pre-processing

(iii) Outliers in dataset

(iv) Decision tree

(v) Association rule mining

(vi) Data mining challenges.