Total number of printed pages—4

53 (IT 701) DTMN

2021

(Held in 2022)

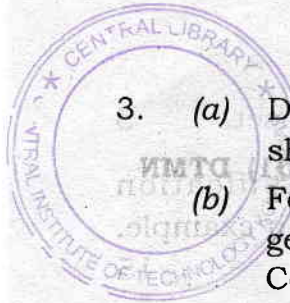**DATA MINING**

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 the term data mining.   2

   (b) Why is pre-processing required in data
       mining ?   3

   (c) With suitable diagram, explain the
       process of knowledge discovery from
       databases.   15

2. Write the steps of K-means clustering
   algorithm. Also state its limitations.

                                    15+5=20

3. *(a)* Discuss Apiori algorithm and its shortcomings. **5**

(b) For the following transaction data set, generate rules using Apriori algorithm. Consider the values as SUPPORT = 50% and Confidence = 75% **15**

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

4. *(a)* What do you mean by frequent pattern (FP) growth algorithm ?

(b) Generate FP-tree for the following transaction data set, which has SUPPORT = 3 :

Transaction ID	Item Purchased
1.	E, A, D, B
2.	D, A, C, E, B
3.	C, A, B, E
4.	B, A, D
5.	D
6.	D, B
7.	A, D, E
8.	B, C

20

5. (a) Define classification of dataset. 5

   (b) Describe the KNN classification algorithm with the help of an example. 15

6. (a) Why the K-means clustering algorithm, divide the following dataset into two clusters? 15

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

   (b) Discuss decision tree-based method for classification. 5

7. Write short notes on : 4×5=20

   (a) Data pre-processing

   (b) Supervised and unsupervised methods

(c) Training dataset and test dataset

(d) Cluster analysis

(e) Datamining challenges

———————