53 (IT 701) DTMN

2021

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)* State the significance of data mining. 8

   (b) Describe the Knowledge Discovery in Databases (KDD) model with flowchart. 12

2. *(a)* List some data mining challenges. Why is data pre-processing required ? 10

   (b) Define classification of data set. Describe the KNN classification algorithm. 10

3. *(a)* Justify, why is clustering of data referred as an unsupervised learning method. 5

   (b) Explain the different approaches of hierarchical data clustering. Discuss the advantages and disadvantages of hierarchical data clustering. 15

4. *(a)* Discuss the Apriori Algorithm. 10

   (b) What is association rule mining? How is decision tree useful in association rule mining? 10

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. Write short notes on the following :
   (any four) 20

    (a) Data Pre-processing

    (b) Supervised and Unsupervised method

    (c) Outlier in data set

    (d) Data mining challenges

    (e) Cluster analysis.