

Total No. of Questions : 8]

SEAT No. :

P3853

[Total No. of Pages : 2

[5561]-281

DATA MINING TECHNIQUES AND APPLICATIONS

B.E. Computer (410444D)

(2012 Pattern) (Semester -I) (End Sem) (Elective-I)

Time : 2½ Hours]

[Max. Marks :70

Instructions to the candidates:

- 1) *Answer Q.1 or Q.2, Q.3 or Q.4, Q.5 or Q.6, Q.7 or Q.8.*
- 2) *Neat diagrams must be drawn wherever necessary.*
- 3) *Figures to the right side indicate full marks.*
- 4) *Assume suitable data if necessary.*

- Q1)** a) Explain any two methods for attribute subset selection. [6]
- b) What are applications of Market basket analysis? What is frequent item set? [6]
- c) Write and explain working of K-NN classification algorithm. [8]

OR

- Q2)** a) Write a note on [8]
- i) Mining multilevel association rules.
 - ii) Constraint based association rule mining.
- b) The data for price is given here: 4, 8, 15, 21, 21, 24, 25, 28, 34. [8]
Apply binning methods to smooth the data.
- c) Compare supervised and Unsupervised learning. [4]
- Q3)** a) What is meant by cluster analysis? [3]
- b) Explain AGNES and DIANA Hierarchical Clustering with example and diagram. [8]
- c) Explain K-medoid clustering algorithm. [6]

OR

P.T.O.

- Q4)** a) How K-Means clustering algorithm is applied in clustering the data? Explain with suitable example. [8]
- b) The two objects are the tuples (22, 1, 42, 10) and (20, 0, 36, 8): [9]
- i) Compute the Euclidean distance between the two objects.
 - ii) Compute the Manhattan distance between the two objects.
 - iii) Compute the Minkowski distance between the two objects.
- Q5)** a) Explain working of web crawlers. [6]
- b) Differentiate between web content mining and web usage mining. [6]
- c) Explain following terms: [5]
- i) Term Frequency
 - ii) Inverse Document Frequency.

OR

- Q6)** a) Explain the following text mining approaches [9]
- i) The tagging approach.
 - ii) The keyword. based approach.
 - iii) The information-extraction approach.
- b) TF-IDF has been used as an effective measure in document classification. [8]
- i) Give one example to show that TF - IDF may not be always a good measure in document classification.
 - ii) Define another measure that may overcome this difficulty.
- Q7)** a) Write a note on advanced techniques for big data mining. [8]
- b) Write note on Reinforcement learning. [8]

OR

- Q8)** a) Draw the diagram for Systematic Machine learning Framework and explain the working. [8]
- b) Write a note on Multi-perspective learning and Holistic Learning. [8]

