[5354]-677

B.E. (Computer Engineering)
DATA MINING TECHNIQUES AND APPLICATIONS
(2012 Pattern) (Semester - I) (Elective - I)

Time : 2½ Hours]   *[Max. Marks : 70*

Instructions to the candidates:

1) *Attempt Q1 or Q2, Q3 or Q4, Q5 or Q6, Q7 or Q8.*
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 Data Preprocessing Steps in brief. **[8]**

b) A database has Nine transactions. Let minimum support count is 2. **[6]**

TID	List of itemIDs
T100	I1,I2,I5
T200	I2,I4
T300	I2,I3
T400	I1,I2,I4
T500	I1,I3
T600	I2,I3
T700	I1,I3
T800	I1,I2,I3,I5
T900	I1,I2,I3

Generate all candidate item set and frequent item sets using Apriori algorithm.

c) Write and explain K-Nearest -Neighbor Classification Algorithm with suitable example. **[6]**

OR

Q2) a) Explain binning methods for data cleaning with suitable example. **[6]**

b) Explain mining Multilevel association rules. What is Uniform support?**[6]**

c) What is decision tree pruning? Explain Error based pruning and Reduced Error Pruning methods for Tree Pruning. **[8]**

P.T.O.

Q3) a) Given two objects represented by the tuples (24, 1, 42, 10) and (20, 0, 36, 8):

     i) Compute the Euclidean distance between the two objects. **[6]**

     ii) Compute the Manhattan distance between the two objects.

  b) What is Hierarchical Clustering? Explain Agglomerative Nesting and Divisive Analysis Hierarchical Clustering methods with example and diagram. **[5]**

  c) Write algorithm for K-means clustering and explain it with suitable example. **[6]**

<div align="center">OR</div>

Q4) a) Write equations for Minimum distance, Maximum Distance and Average Distance Formulas for measuring distance between clusters. Explain nearest neighbour clustering, single linkages algorithm farthest neighbour clustering algorithm, complete linkages algorithm with help of these formulas. **[6]**

  b) Explain K- Medoids clustering algorithm with example. **[6]**

  c) Explain Partitioning methods in large databases. **[5]**

Q5) a) What is a dimensionality reduction technique in text mining? List all techniques and explain any one of them. **[8]**

  b) Explain Web structure Mining and web content mining. **[6]**

  c) Explain Precision and Recall in the terms of retrieved and relevant documents. **[3]**

<div align="center">OR</div>

Q6) a) Write Crawling Algorithm and explain working of Web Crawler. **[8]**

  b) Explain following terms: **[9]**

     i) TF

     ii) IDF

     iii) Feature Vector

Q7) a) Write short note on **[6]**

      i) Holistic learning

      ii) Systematic learning.

   b) What is Big Data? Explain in short. What are the challenges in Big data Analysis? **[4]**

   c) What is supervised, unsupervised and semi supervised learning? **[6]**

OR

Q8) a) Write a note on Advances in technology for big data mining. **[8]**

   b) Draw a diagram for reinforcement learning model and explain it in detail. **[8]**

◯◯◯