

Total No. of Questions : 8]

SEAT No. :

P2042

[Total No. of Pages : 3

[5059]-647

B.E. (Computer Engineering)

**DATA MINING TECHNIQUES AND APPLICATIONS**  
**(2012 Pattern) (Semester - I)**

*Time : 2½ Hours]*

*[Max. Marks : 70*

*Instructions to the candidates:*

- 1) Answer 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) In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem.

[6]

b) Explain the following terms :

[6]

- i) Constraint-based rule mining.
- ii) Closed and maximal frequent itemsets.

c) Consider the following data for a binary class problem

[8]

A	B	Class
T	F	P
T	T	P
T	T	N
T	F	P
T	T	P
F	F	N
F	F	N
F	F	N
T	T	P
T	F	N

**P.T.O.**

- i) Compute the information gain for A1 and A2.
- ii) What is the best split between A1 and A2 according to Information gain?
- iii) Compute the Gini index for A1 and A2.
- iv) What is the best split between A1 and A2 according to Gini index?

OR

**Q2)** a) Consider the market basket transactions shown below : **[6]**

Transaction ID	Items bought
T1	{M, A, B, D}
T2	{A, D, C, B, F}
T3	{A, C, B, F}
T4	{A, B, D}

Assuming the minimum support of 50% and minimum confidence of 80%

- i) Find all frequent itemsets using Apriori algorithm.
  - ii) Find all association rules using Apriori algorithm
- b) What are the major tasks in data preprocessing? Explain them in brief. **[6]**
- c) Explain with suitable example : **[8]**
- i) k-Nearest-Neighbor Classifier
  - ii) Scalable decision tree

**Q3)** a) Consider the following points six points : **[8]**

P1(0.40, 0.53), P2(0.22, 0.38), P3(0.35, 0.32), P4 (0.26, 0.19), P5(0.08, 0.41) and P6(0.45, 0.30).

Perform the single link hierarchical clustering and show your results by drawing a dendrogram.

- b) Explain with suitable example the k-medoids algorithm **[6]**
- c) What are the requirements of clustering in data mining? **[3]**

OR

**Q4)** a) What is meant by cluster analysis? **[4]**

b) Explain with suitable example the K-means algorithm. **[5]**

- c) Differentiate between following clustering methods **[8]**
- i) Single and complete link
  - ii) Hierarchical and partitioning

- Q5)** a) Precision and recall are two essential quality measures of an information retrieval system. [6]  
i) Why it is usual practice to trade one measure for the other? Explain.  
ii) Why F-score is a good measure for trade between precision and recall.
- b) Compare the different text mining approaches. [5]
- c) Explain the following terms : [6]  
i) Bag of words  
ii) Feature vector

OR

- Q6)** a) What is Web usage mining? Explain in brief. [6]  
b) Differentiate between document selection and document ranking methods of information retrieval. [5]  
c) Explain the following terms : [6]  
i) Authoritative Web pages  
ii) Hub pages  
iii) Document Object Model (DOM) structure
- Q7)** a) What is meant by machine learning? Differentiate between supervised and unsupervised machine learning. [6]  
b) What are the similarities and differences between reinforcement learning and artificial intelligence algorithms? [5]  
c) Write short note on mining of big data. [5]

OR

- Q8)** a) What is meant by wholistic learning? [4]  
b) Briefly explain the reinforcement learning. [6]  
c) What is meant by multi-perspective decision making? Explain. [6]



