

Total No. of Questions :10]

SEAT No. :

**P3428**

**[5670] -704**

[Total No. of Pages :3

**B.E. (Computer Engg.)**

**DATA MINING AND WAREHOUSING**

**(2015 Pattern) (410244D) (Semester-I) (End Sem.)(Elective-I)**

*Time :2½Hrs]*

*[Max. Marks : 70*

*Instructions to the candidates*

- 1) *Assume suitable data, if necessary.*
- 2) *Neat diagrams must be drawn wherever necessary.*
- 3) *Figures to the right indicate full marks.*

- Q1) a)** How to compute dissimilarity for nominal attributes with examples. [4]
- b)** Explain Knowledge discovery from data or KDD. [6]

OR

- Q2) a)** Explain data matrix and data dissimilarity matrix with example. [4]
- b)** Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 16, 17, 20, 22, 25, 25, 28, 30, 35, 40, 42, 42, 45, 50, 52, 55, 60, 60, 62, 62, 65, 70, 72 partition them into three bins by each of the following methods. [6]
- i) equal-frequency partitioning
  - ii) equal-width partitioning
- Use smoothing by bin means to smooth the above data.*

- Q3) a)** Explain attribute subset selection techniques. [4]
- b)** Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. draw a schema diagram for the above data warehouse [6]

OR

**P.T.O.**

- Q4) a)** Explain following. [4]  
 i) Manhattan distance  
 ii) Supremum distance  
 b) Briefly compare the following concepts. You may use an example to Explain your point(s). Snowflake schema, fact constellation, starlet query model [6]

- Q5) a)** Consider the market basket transactions shown below: [8]

TransactionID	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) Explain mining Multilevel association rules. What is Uniform suport?[6]  
 c) Define closed frequent itemset and maximal frequent itemset. [2]

OR

- Q6) a)** A database has five transactions. Let  $min\_sup = 60\%$  and  $min\_conf = 80\%$ . [8]

TID	items-bought
T100	{M,O,N,K,E,Y}
T200	{D,O,N,K,E,Y}
T300	{M,A,K,E}
T400	{M,U,C,K,Y}
T500	{C,O,O,K,I,E}

- i) Find all frequent itemsets using FP-growth  
 ii) List all the strong association rules (with support  $s$  and confidence  $c$ )

- b) Explain the techniques to improve the efficiency of apriori algorithm. [6]  
 c) Explain strong association rule with example. [2]
- Q7)** a) Define classification and prediction. Explain decision tree based classification method with suitable example. [8]  
 b) Describe K-Nearest Neighbor classifiers with suitable example. [6]  
 c) Write short note on Rule Induction Using a sequential Covering Algorithm. [4]
- OR
- Q8)** a) Explain the following with example. [8]  
 i) Gini index  
 ii) Entropy  
 iii) Information gain  
 b) Differentiate between supervised and unsupervised Learning [6]  
 c) Explain case based reasoning classifier. [4]
- Q9)** a) Explain following with example [8]  
 i) Accuracy  
 ii) precision  
 iii) Recall  
 iv) Specificity  
 b) Describe following. [8]  
 i) Multiclass classification  
 ii) Reinforcement learning
- OR
- Q10)** a) Explain in detail following techniques to evaluate the accuracy of a Classifier [8]  
 i) Random subsampling  
 ii) Cross validation  
 b) Explain following. [8]  
 i) Systematic learning  
 ii) Wholistic learning

