

Time: 3 hours

Max. Marks: 80

**N.B. (1) Question one is Compulsory.****(2) Attempt any 3 questions out of the remaining.****(3) Assume suitable data if required.**

- Q. 1 (a) Explain data warehouse features. 05  
 (b) Demonstrate with diagram the process of KDD. 05  
 (c) What is Market basket analysis? 05  
 (d) Explain with example confusion matrix, accuracy and precision. 05

- Q. 2 (a) (i) Explain with example Star schema and Snowflake schema. 10  
 (ii) Explain with example any four OLAP operations.  
 b) What is clustering? Explain K-mean clustering algorithm. Suppose that the data mining task is to cluster the following items into two clusters.  
 {4, 8, 20, 24, 6, 40, 60, 22, 50}. Apply k-means algorithm. 10

- Q.3 a) i) Data preprocessing is necessary before data mining process". Justify your answer. 05  
 ii) Explain any 2 data normalization techniques. 05  
 b) i) Explain with example Support, Confidence and Lift function in association mining. 05  
 ii) Consider the transaction database given below:  
 Use Apriori Algorithm with min-support = 50% and min-confidence = 60%,  
 to find frequent itemset and strong association rules. 05

TID	Items
100	a,c,d
200	b,c,e
300	a,b,e
400	b,e

- Q. 4 a) Illustrate any one classification technique for the following dataset.  
 Show how we can classify new tuple (**Homeowner=YES, Status =Employed, Income = Average**). 10

Sr. No	Homeowner	Status	Income	Defaulted
1	Yes	Employed	High	No
2	No	Business	Average	No
3	No	Employed	Low	No
4	Yes	Business	High	No
5	No	Unemployed	Average	Yes
6	No	Business	Low	No
7	Yes	Unemployed	High	No
8	No	Employed	Average	Yes
9	No	Business	Low	No
10	No	Employed	Average	Yes

- b) What is web mining? Explain web content mining in detail 10
- Q. 5 a) Explain FP tree with appropriate example. 10  
 b) Clearly explain the working of DBSCAN algorithm using appropriate diagram 10
- Q. 6 a) Explain Multidimensional and multilevel rule mining with example. 10  
 b) Explain with example different data sampling techniques. 10