

Time: 3 Hrs

Max Marks: 80

N.B.: (1) Question No. 1 is compulsory.

(2) Attempt any **three** questions out of remaining **five**.

1. (a) Apply K-means Algorithm to divide the given set of values {2,3,6,8,9,12,15,18,22} into 3 clusters . 5

- (b) Explain Confusion Matrix. Calculate Accuracy, Precision and Recall for the following Confusion Matrix 5

Cancer Classes	Yes	No	Total
Yes	90	210	300
No	140	9560	9700
Total	230	9770	10000

- (c) What are the major issues in data mining? 5

- (d) Explain different types of attributes in Data Mining. 5

2. (a) Consider the following transaction database: 10

TID	Items
01	A, B, D,E,F
02	B,C,E
03	A, B, D, E
04	A,B,C,E
05	A,B,C,D,E,F
06	B,C,D
07	A,B,D,E

Apply the Apriori algorithm with minimum support of 30% and minimum confidence of 75% and find the association rules in the data set.

- (b) What is linear regression? Explain with example. 10

3. (a) Explain BIRCH algorithm with example. 10

- (b) What are outliers? Explain different outlier detection techniques. 10

4. (a) Using the given training dataset classify the following tuple using Naïve Bayes Algorithm: <Homeowner: No, Marital Status: Married, Job experience:3> **10**

Homeowner	Marital Status	Job experience (in years)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

- (b) Define Business intelligence and decision support systems with examples. **10**
5. (a) What is Data Mining? Explain KDD process with diagram. **10**
- (b) Explain Market-Basket Analysis with example. **10**
6. (a) What are multiple level and multidimensional association rules? Explain with suitable examples for each. **10**
- (b) Suppose that data for analysis includes the attribute age. The age values for data tuples are **10**
 13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35, 36, 40,45,46,52,70
 i) What is mean of the data? What is median of data?
 ii) What is mode of data? Comment on data's modality.
 iii) What is mid-range of data?
 iv) Give the five-point summary of the data.
 v) Show box plot of the data.
