

(3 Hours)

[Total Marks:80]

- NB:** 1. Question no. 1 is compulsory.  
 2. Answer any **three** out of the **remaining** questions.  
 3. Assume data, if missing, with justification.

**Q.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 [05]

**(b)** Explain Confusion Matrix. Calculate Accuracy, Precision and Recall for the following Confusion Matrix [05]

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? [05]

**(d)** What is noisy data? How to handle it? [05]

**Q.2.(a)** Consider the transaction database given in table below. Apply Apriori Algorithm with minimum support of 50% and confidence of 50%. Find all frequent itemsets and all the association rules. [10]

Tid	Items
100	1,3,4
200	2,3,5
300	1,2,3,5
400	2,5
500	1,2,3
600	3,5
700	1,2,3,5
800	1,5
900	1,3

**(b)** Explain Regression. Explain linear regression with example. [10]

**Q.3.(a)** Suppose we have five objects with name A, B, C, D and E. Apply single linkage clustering and draw dendrogram for the given data. [10]

	X	Y
A	1	1
B	1.5	1.5
C	5	5
D	3	4
E	4	4
F	3	3.5

**(b)** What is an outlier? Describe methods that are used for outlier analysis. [10]

**Q4.(a)** Using the given training dataset classify the following tuple using Naïve Bayes Algorithm: [10]  
 <Homeowner: No, Marital Status: Married, Job experience:3>

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)** Explain Business Intelligence issues. [10]

**Q5. (a)** What is data mining? Explain KDD process with diagram. [10]

**(b)** Explain Market-Basket analysis with example. [10]

**Q6. (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 (in increasing order): [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 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.

\*\*\*\*\*