Duration: 3hrs **[Max Marks:80]**

N.B. : (1) Question No 1 is Compulsory.
(2) Attempt any three questions out of the remaining five.
(3) All questions carry equal marks.
(4) Assume suitable data, if required and state it clearly.

Q.1 [20]

A Give the difference between OLAP and OLTP

B Explain Confusion Matrix. Calculate Accuracy, Precision and Recall for the following Confusion Matrix

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

C Explain Multilevel and Multidimensional Association rules with suitable examples

D Explain DBSCAN algorithm with example.

Q.2 A What is noisy data? How to handle it [10]
For the following data D={4,8,9,15,21,21,24,25,26,28,29,34}
Number of bins =3
Perform the following:
i. Partition into equal frequency bins
ii. Smoothing by bin means
iii. Smoothing by bin boundaries

B Explain K means algorithm in detail. Apply K-means Algorithm to divide the given set of values {2,3,6,8,9,12,15,18,22} into 3 clusters [10]

Q.3 A Explain different types of attributes used in data exploration with example [10]

B 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

Q.4 A What do you mean by data mining? Explain KDD process with help of a suitable diagram [10]

B For the table given below apply Apriori algorithm and show frequent item set and strongassociation rules. Assume Minimum Support of 30% and Minimum confidence of 70%. [10]

TID	Items
01	1,3,4,6
02	2,3,5,7
03	1,2,3,5,8
04	2,5,9,10
05	1,4

Q.5 A Design BI system for Fraud Detection? Explain all steps from data collection to decision making [10]

B Compare star schema, Snow flakes schema. [10]

Q.6 **Solve the following(Any 2)** [20]

A What is an outlier? List types of outliers. Describe methods used for outlier analysis.

B Describe different steps involved in data preprocessing

C Explain Bagging of a classifier

****