

Time: 03 Hours

Marks: 80

**Note:** 1. Question 1 is compulsory

2. Answer any three out of remaining questions.

- Q1 A) i. Design star & snowflake schema for "Hotel Occupancy" considering [10] dimensions like Time, Hotel, Room, etc.  
 ii. Calculate the maximum number of base fact table records for the values given below:  
 Time period: 5 years  
 Hotels: 150  
 Rooms: 750 rooms in each Hotel (about 400 occupied in each hotel daily).

B) Explain Data mining as a step in KDD. Give the architecture of typical data mining [10] System.

- Q2 A) The college wants to record the marks for the courses completed by students using [10] the dimensions: a) Course, b) Student, c) Time & a measure d) Aggregate marks. Create a Cube and describe following OLAP operations:

i) Rollup ii) Drill down iii) Slice iv) Dice v) Pivot.

B) A simple example from the stock market involving only discrete ranges has profit [10] as categorical attribute, with values {up, down} and the training data is:

Age	Competition	Type	Profit
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	Up

Apply decision tree algorithm and show the generated rules.

- Q3 A) Why naive Bayesian classification is called "naive"? Briefly outline the major ideas [10] of Naive Bayesian classification.

B) Discuss different steps involved in Data Pre-processing [10]

- Q4 A) Explain ETL of data warehousing in detail. [10]

B) Find clusters using  $k$ -means clustering algorithm if we have several objects [10] (4 types of medicines) and each object have two attributes or features as shown in the table below. The goal is to group these objects into  $k=2$  group of medicine

based on the two features (pH and weight index).

Object	Attribute 1(X) Weight Index	Attribute 2 (Y) pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

Q5 A) Explain Data Warehouse Architecture in detail. [10]

- B) A database has five transactions. Let minimum support = 30% and minimum Confidence = 70% [10]
- Find all frequent patterns using Apriori Algorithm.
  - List strong association rules.

Transaction_Id	Items
A	1,3,4,6
B	2,3,5,7
C	1,2,3,5,8
D	2,5,9,10
E	1,4

Q6 Write short note on the following (Answer any **FOUR**) [20]

- Data warehouse design strategies
- Applications of Data Mining
- Role of metadata
- Multidimensional and multilevel association mining
- Hierarchical clustering

\*\*\*\*\*