Data Warehouse And Mining - May 18
Computer Engineering (Semester 8)

Total marks: 80
Total time: 3 Hours

INSTRUCTIONS
(1) Question 1 is compulsory.
*(2) Attempt any **three** from the remaining questions.*
(3) Draw neat diagrams wherever necessary.

1.a.i. Design star & snowflake schema for "Hotel Occupancy" considering dimensions like Time, Hotel, Room, etc. (10 marks)

1.a.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). (5 marks)

1.b. Explain Data mining as a step in KDD. Give the architecture of typical data mining System. (10 marks)

2.a. The college wants to record the marks for the courses completed by students using 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. (10 marks)

2.b. A simple example from the stock market involving only discrete ranges has profit 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

Age	Competition	Type	Profit
New	No	Hardware	Up
New	No	Software	Up

Apply decision tree algorithm and show the generated rules. (10 marks)

3.a. Why naive Bayesian classification is called "naive"? Briefly outline the major ideas of Naive Bayesian classification. (10 marks)

3.b. Discuss different steps involved in Data Pre-processing. (10 marks)

4.a. Explain ETL of data warehousing in detail. (10 marks)

4.b. Find clusters using k-means clustering algorithm if we have several objects (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). (10 marks)

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

5.a. Explain Data Warehouse Architecture in detail. (10 marks)

5.b. A database has five transactions. Let minimum support = 30% and minimum confidence = 70%

i. Find all frequent patterns using Apriori Algorithm. (10 marks)

ii. 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

Q.6 Write short note on the following (Answer any FOUR)

6.a. Data warehouse design strategies   (5 marks)
6.b. Applications of Data Mining   (5 marks)
6.c. Role of metadata   (5 marks)
6.d. Multidimensional and multilevel association mining   (5 marks)
6.e. Hierarchical clustering   (5 marks)