

1. What is data mining? Explain the significance of data mining in present day business scenario. [2+6]
2. List the different stages of Knowledge Discovery in Database. Explain the major task performed in each stage, with example. [10]
3. Discuss the concept of frequent sets, confidence and support. Distinguish supervised classification and unsupervised classification technique. [6+2]
4. What is clustering? Differentiate between the partitioning clustering and hierarchical clustering technique. Are the outliers desirable while performing cluster analysis? [3+4+3]
5. What is data warehouse? List and explain the characteristics of data warehouse. “Within a data warehouse, the fact tables are never updated, however, dimension tables are supposed to be updated frequently”. Do you agree? Make your comment. [2+4+3]
6. Discuss the role of load manager, warehouse manager and query manager within the data warehouse. Differentiate between the operational database and data warehouse database. [6+3]
7. Why does an organization need Online Analytical Processing? Explain the different operations associated with OLAP. [8]
8. An insurance company has customers with the attributes as shown in the table below. The company wants to identify the two individuals that have the behavior similar to Customer-1. Use likelihood and distance approach to identify those two customers.

| Customer | Age | Income | Credit |
|-----------------|------------|---------------|---------------|
| Customer-1 | 23 | 130,000 | 0 |
| Customer-2 | 35 | 235,000 | 535,000 |
| Customer-3 | 25 | 105,000 | 50,000 |
| Customer-4 | 30 | 275,000 | 455,000 |
| Customer-5 | 45 | 375,000 | 300,000 |
| Customer-6 | 49 | 315,000 | 270,000 |

[6]

9. XYZ Departmental Store is a popular departmental store in the New Road. The departmental store sales the following set of items

| | | |
|--------|--------|-------|
| Peanut | Butter | Bread |
| Coke | Soda | Jam |

The manager of the departmental store wants to know the relationships between sales of the above items. You, as an expert in data mining, are given a following transaction database. Deduce some rules that the manager could use in his

business. Use Apriori algorithm. (Assume minimum support=20% and confidence=65%).

| Transac ID | Items |
|-------------------|-----------------------------|
| 1 | Peanut, Butter, Bread |
| 2 | Butter, Bread, Coke |
| 3 | Peanut, Bread, Coke |
| 4 | Bread, Coke, Soda |
| 5 | Peanut, Butter, Bread, Coke |
| 6 | Peanut, Bread, Coke, Jam |
| 7 | Peanut, Butter, Bread |
| 8 | Butter, Coke |

[12]

