

(3 Hours)

80 Marks

N.B.: (1) Q.1 is compulsory.
 (2) Attempt any three out of remaining five.
 (3) Figures to the right indicate full marks.

- 1 A) Discuss issues to consider during data integration. [10]
- B) Describe in detail Data warehouse architecture and ETL process. [10]
- 2 A) What is **Market Basket Analysis**? Find out strong association rule form the given example using apriori algorithm with the support of 50% and confidence of 70% [10]

Trans_Id	Item
1.	Laptop, Mouse, Headphones, Pendrive, Speakers
2.	Laptop, Headphones
3.	Laptop, Mouse, Pendrive
4.	Mouse, Speakers
5.	Laptop, Pendrive

- B) Explain text mining and discus in brief the information retrieval methods. [10]
- 3 A) What are the characteristics and benefits of data marts? [10]
- B) Explain data pre-processing in detail. Apply Naïve Bayes algorithm and predict that if a fruit has the following properties then which type of fruit it is. [10]
 Fruit{ Yellow, Sweet, Long }

Fruit	Yellow	Sweet	Long	Total
Mango	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

- 4 A) What is classification? Explain in detail Associative Classification method. [10]
- B) Write the algorithm for K-Means Clustering. Generate the clusters using K-Means Clustering (K = 2) [10]

Object	Attribute1 (X): weight index	Attribute 2 (Y): pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

- 5A) Illustrate, with an example, the following OLAP operations: roll-up, drill-down, slice, dice, and pivot. [10]
- B) What is regression? How can linear regression be used for prediction. [10]

- 6 A) Write Notes on any two: [10]
 - a) Web mining.
 - b) Star and snow flake schema
 - c) KDD process.

- B) Explain the prediction methods and models for business intelligence. [10]
