

N.B

- 1) All questions are compulsory.
- 2) Figures to the right indicate marks.
- 3) Illustrations, in-depth answers and diagrams will be appreciated.
- 4) Mixing of sub-questions is not allowed.
- 5) Use of simple calculator is allowed.

Q.1 Attempt any three.

[15 Marks]

- a. Explain Data Warehouse Architecture with suitable diagram.
- b. Why Metadata is important? What are the different types of metadata?
- c. What is Data Warehouse? What are different features of it?
- d. Explain Operational Versus Decision-Support Systems.
- e. What is Data Mart? Explain with suitable example.
- f. Write short note on Web-Enabled Data Warehouse.

Q.2 Attempt any three.

[15 Marks]

- a. What is Dimension modeling? Explain benefits of dimension modeling.
- b. Write short note on dimension and fact table.
- c. What is need of Data design?
- d. What is Snowflakes schema? Explain with suitable example.
- e. What is ETL? Explain different steps for ETL process.
- f. What is Star schema? Explain with suitable example.

Q.3 Attempt any three.

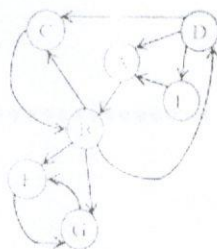
[15 Marks]

- a. What are the difference between Data Mining and Machine Learning?
- b. What Can Data Mining Do and Not Do?
- c. What is Classification? What are different types of Classification?
- d. What are different Data Mining Techniques?
- e. Write a short note on Weka.
- f. Explain Naive Bayes Method with suitable example.

Q.4 Attempt any three.

[15 Marks]

- a. Explain different characteristics of Clustering.
- b. What are different applications of Cluster Analysis?
- c. Write a short note on k-means clustering.
- d. What is Manhattan distance? Consider three person located at (30,70), (40,54) and (80,50) respectively calculate similarity distance among person using Manhattan distance.
- e. What is Page Rank algorithm? Calculate page rank of following web pages with damping factor (d) is 0.7.



- f. Explain the architecture of search engine with suitable diagram.

Q.5 Attempt any three.

[15 Marks]

- What are the different approaches for Transaction Database Storage?
- Write a short note on Association Rule Mining.
- What is the FP-Growth algorithm?
- Find the association rules having minimum support of 50% and minimum confidence of 75%.

TID	ITEMS
1	Butter, Bread
2	Butter, Bread, Jam
3	Butter, Eggs
4	Bread, Jam, Eggs

- Explain Apriori Algorithm.
- Consider the sales database given below.

TID	ITEMS
1	Pen, Scale
2	Pencil, Eraser, Sharpener
3	Book, Notebook, Scale, Sharpener
4	Pen, Pencil, Eraser, Scale, Book

Calculate:

- Support(Pencil)
- Support(Pen, Scale)
- Confidence(Pencil, Eraser)
- Confidence(Book, Scale)
- Lift(Book, Scale)
