
Question Paper consists of FIVE units, each carrying 14 marks
Each unit has TWO questions; either of them should be answered
All parts of a question must be answered at one place.

UNIT-I

1. a) What is KDD? Explain about data mining as a step in the process of knowledge discovery. 7M
b) Explain the differences between structured, semi-structured, and unstructured data. 7M

(OR)

2. a) What are the various types of OLAP servers? Explain. 7M
b) Discuss about a three-tier data warehouse architecture. 7M

UNIT-II

3. a) What is entropy and information gain in the context of decision trees? 7M
b) Discuss the k-Nearest Neighbour (k-NN) algorithm with suitable example. 7M

(OR)

4. a) Explain the concept of Naive Bayes classifier with example. 7M
b) What is the difference between linearly separable and non-separable data? How does SVM handle non-linearly separable data? 7M

UNIT-III

5. a) What is the difference between a frequent itemset, a closed frequent itemset, and a maximal frequent itemset? 7M
b) Explain about the Apriori algorithm for finding frequent item sets with an example. 7M

(OR)

6. a) Illustrate FP-growth algorithm with a suitable example. 7M
b) What are subgraph patterns? Give an example from bioinformatics or social networks. 7M

UNIT-IV

7. a) Define Clustering? Explain about Types of Data in Cluster Analysis? 7M
b) Discuss the K Mean algorithm with suitable example. 7M

(OR)

8. a) Explain the DBSCAN with suitable example. 7M
b) Elaborate Unsupervised Cluster Evaluation using cohesion and separation. 7M

UNIT-V

9. a) What is web data mining? How is it different from traditional data mining? 7M
b) Describe the steps in web usage mining: preprocessing, pattern discovery, and pattern analysis. 7M

(OR)

10. a) Write are the key characteristics of web data? Why is web data considered semi-structured or unstructured? 7M
b) Write Short notes on Search Engines. 7M