

DATA WAREHOUSING AND DATA MINING

(Information Technology)

Time: 3 Hours

Max. Marks: 70

Answer any FIVE Questions
All Questions carry equal marks

1. What is the need for preprocessing the data? Explain briefly various forms of data preprocessing.

2. (a) Discuss multidimensional data model and explain various schemes for multidimensional data model.

(b) Explain indexing OLAP data.

(c) Describe types of OLAP servers.

3. (a) Find frequent itemsets for the following table using FP-Growth algorithm. Assume relevant thresholds.

T _{id}	List of item ids
T ₁	I ₁ , I ₃ , I ₅
T ₂	I ₂ , I ₄ , I ₁
T ₃	I ₁ , I ₂ , I ₃ , I ₄
T ₄	I ₅ , I ₃ , I ₂
T ₅	I ₁ , I ₂ , I ₅
T ₆	I ₃ , I ₄ , I ₅

- (b) Discuss constraint based mining.

4. (a) State Baye's theorem. Explain how it can be adopted for classification.

(b) Describe case-based reasoning as a lazy learner.

(c) List the measures for classifier's accuracy.

5. (a) Discuss hierarchical methods for clustering. List their merits and demerits.

(b) Explain statistical based outlier detection.

6. Discuss the characteristics of social network and the tasks challenges in mining social networks.

7. (a) Write about probabilistic latent semantic indexing method.

(b) Explain HITS algorithm.

(c) What is meant by an object cube?

8. (a) Discuss data mining for financial data analysis in brief.

(b) Write a note on statistical data mining.