

No. of Printed Pages : 3

RCSE-001

**Ph. D. (COMPUTER SCIENCE)
(PHDCS)**

**Term-End Examination
December, 2025**

RCSE-001 : DATA MINING

Time : 3 Hours

Maximum Marks : 100

Weightage : 50%

Note : (i) *Question No. 1 is compulsory.*

(ii) *Answer any **three** questions from the rest.*

-
-
1. (a) What is Information Package Diagram (IPD) in Data Warehousing ? What is its purpose and what are its contents ? Illustrate with the help of an example. Convert the IPD considered in your example into star-schema. 10

- (b) What is OLAP ? Draw and explain MOLAP, ROLAP and HOLAP architectures. 10
- (c) State Bayes' theorem. How can it be applied for data classification. Explain with an example. 5
- (d) With example explain Bayesian-Belief Network. 5
- (e) Explain the Apriori algorithm for finding frequent item sets with an example. 10
2. (a) What is ETL in Data Warehousing ? Explain the working of ETL with a block diagram. 10
- (b) Along with suitable block diagrams, explain the 2-tier and 3-tier data warehouse architectures. 10
3. (a) What are Association rules ? Explain the process of generating Association rules from frequent itemsets with appropriate examples. 10

[3]

- (b) Explain Rule-based classification. With an example, explain how the Rule Extraction is performed from a Decision Tree ? 10
4. (a) What is data-preprocessing ? Why is it important ? Explain all the key-steps in data preprocessing. 10
- (b) Explain in detail about attribute selection measures in classification. 10
5. Write short notes on the following : $4 \times 5 = 20$
- (a) Mining frequent item-sets using vertical data formats
- (b) K-nearest Neighbour Classifiers
- (c) Ensemble methods
- (d) Fact constellations

× × × × ×