No. of Printed Pages : 4 **RCSE–001**

Ph. D. IN COMPUTER SCIENCE (PHDCS)
Term-End Examination
June, 2026

RCSE-001 : DATA MINING

Time : 3 Hours *Maximum Marks : 100*

Note : *(i) Question No. **1** is compulsory.*
*(ii) Answer any **three** questions from the rest.*

1. (a) Critically analyse the evolution of Data Warehousing and discuss how the approaches of Bill Inmon and Ralph Kimball shaped modern Data Warehouse design philosophies. 10

   (b) Assess the impact of Cloud-based Data Warehouse Architectures on scalability, cost-effectiveness and data accessibility in enterprise environments. 10

(c) Explain the concept of frequent pattern mining and association rule mining in data mining with suitable example. 10

(d) What is Cluster Analysis ? How is this used in Data Mining ? Also mention few applications of cluster analysis in data mining. 10

2. (a) "Data Marts are both an opportunity and a challenge in enterprise data architecture." Examine this statement with reference to dependent, independent and hybrid data marts. 10

(b) Diffrentiate among ROLAP, MOLAP, HOLAP and DOLAP architectures. Discuss how these impact data storage efficiency, query performance and scalability. 10

3. (a) Discuss the role of Hadoop in supporting Data Warehousing (DW) design and operations. Explain how its ecosystem (MapReduse, Hive, Pig, Sqoop etc.) enhances scalability and performance. 10

(b) Discuss the role of ETL process in ensuring data quality and consistency within a Data Warehouse. 10

4. (a) A database has transactions. Let min_ sup = 60% and min_conf = 80% : 10

TID	Items_bought
T_{100}	{M, O, N, K, E, Y}
T_{200}	{D, O, N, K, E, Y}
T_{300}	{M, A, K, E}
T_{400}	{M, U, C, K, Y}
T_{500}	{C, O, O, K, I, E}

(i) Find all frequent itemsets using FP-growth algorithm.

(ii) List 5 strong association rules (with support s and confidence c).

(b) Discuss the importance of tree pruning in preventing overfitting, and illustrate with an example how pruning affects classification accuracy. 10

5. (a) What are the fundamental metrics used to evaluate the performance of a classification model ? Explain. 10

   (b) Explain the working of K-means algorithm with the help of suitable example. 10

× × × × ×