

No. of Printed Pages : 2

**RCSE–001**

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

**Term-End Examination**

**December, 2023**

**RCSE-001 : DATA MINING**

*Time : 3 Hours*

*Maximum Marks : 100*

*Weightage : 50%*

---

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

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

---

---

1. (a) Explain the evolution of Database System Technology with the help of a diagram. 10
- (b) What kinds of Data that can be mined ? 10
- (c) Suppose that the data for analysis includes the attribute 'age'. The 'age' values for the data tuples are (in increasing order) :  
13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25,  
30, 33, 33, 35, 35, 36, 40, 45, 46, 52, 70
  - (i) What is the mean of the data ? 5
  - (ii) What is mode of the data ? Comment on the data's modality (i.e., limited, trimodal etc.) 5

**P. T. O.**

- (d) *Data Quality* can be assessed in terms of several issues, including accuracy, completeness and consistency. For each of the above three issues, discuss how data quality assessment can depend on *intended use* of the data, giving examples. 10
2. (a) Explain the three tier data warehousing architecture with the help of a diagram. 10
- (b) What are OLTP and OLAP systems ? Distinguish between them. 10
3. (a) Explain any *one* of the Data Cube computation methods. 10
- (b) What are sampling cubes ? What are their advantages ? 10
4. (a) Give an example to show that items in a strong association rule actually may be negatively correlated. 10
- (b) Explain any *one* method for mining in multilevel, multidimensional space. 10
5. (a) What is meant by Constraint-Based Frequent Pattern Mining ? 10
- (b) Explain the process of mining compressed patterns by pattern clustering. 10