

No. of Printed Pages : 4

RCSE-001

**DOCTOR OF PHILOSOPHY IN
COMPUTER SCIENCE (PHDCS)**

Term-End Examination

December, 2021

RCSE-001 : DATA MINING

Time : 3 Hours

Maximum Marks : 100

Weightage : 50%

Note : (i) *Question No. 1 is compulsory and carries 40 marks.*

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

1. (a) What is Data Mining ? Why is data mining not a business solution ? Explain with an example. 10

- (b) What are Support Vector Machines (SVM) ? Discuss the linear SVM case with the help of suitable example. 10

- (c) Define a Data Mart (DM). How are Data Marts different from Data Warehouse ? Draw the architecture diagram of data marts with enterprise data warehouse and explain its components. 10

- (d) Explain the following tasks in data preprocessing :

- (i) Data Cleaning 4
(ii) Dimensionality reduction 3
(iii) Numerosity reduction 3

2. Explain the following architectures :

- (i) ROLAP 8
(ii) MOLAP 7
(iii) Real time analytical processing 5

P. T. O.

[3]

RCSE-001

3. (a) Explain basic requirements of data mining tools. Why framing of relevant questions in data mining is important ? Describe with the help of an example. 10
- (b) Why is “Tree Pruning” useful in decision tree induction ? What is a drawback of using a separate set of tuples to evaluate pruning ? 10
4. (a) 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 the intended use of the data, giving examples. 10
- (b) Explain various issues that needs to be considered during data integration. 10

[4]

RCSE-001

5. Write short notes on the following :
- (i) Laplacian correlation and its utility 4
- (ii) Naive Bayesian classification 4
- (iii) Genetic algorithms 4
- (iv) Classifier accuracy measures 4
- (v) ROC curves 4

RCSE-001

P. T. O.