

PG 2nd Semester Examination, 2022 (R & B)

Subject : ITM

(Data Mining and Data Warehousing)

Paper – 1.2.2

Full Marks : 80

Time : 3 hours

Answer all questions

The figures in the right-hand margin indicate marks

Candidates are required to give their answers in their own words as far as practicable

1. What is the importance of data pre-processing ?
Explain the major tasks of data cleaning. 16

Or

- (a) What is data mining ? What are the major issues in data mining ? 8
- (b) Briefly explain some applications of data mining. 8

(Turn Over)

2. What are the implementations of data warehouse ?
Explain it. 16

Or

- (a) What is data warehouse ? What are the benefits and goals of the data warehouse ? Differentiate between OLTP and OLAP. 8
- (b) Briefly explain Three-Tier data warehousing architecture with suitable diagram. 8

3. What is association rule mining ? 16

For the following given transaction data set, find out frequent set and also derive association rule using Apriori Algorithm. Consider the values as $\text{min_support} = 50\%$ and $\text{min_confidence} = 75\%$

Transaction ID	Items Purchased
1	Bread, Cheese, Egg, Juice
2	Bread, Cheese, Juice
3	Bread, Milk, Yogurt
4	Bread, Juice, Milk
5	Cheese, Juice, Milk

Handwritten notes:
A, B, C
AB → C
BC → A
AC → B
A

Or

- (a) How to find correlation analysis from association mining ? 8
- (b) What is a quantitative association rule ? 8
4. Write short notes on : (Any two) 8 × 2
- (i) Bayesian Classification
 - (ii) Support Vector Machine
 - (iii) Linear Regression
 - (iv) Classifier Accuracy.
5. (a) What are the minimal requirements for domain knowledge to determine input parameters ? 16

Or

What do you mean by clustering ? Divide the following sample data into two clusters using K-means method.

Serial No.	Height (H)	Weight (W)
1	185	72
2	170	56
3	168	60
4	179	68
5	182	72
6	188	77
7	180	71
8	180	70
9	183	84
10	180	88
11	180	67
12	177	76