REG.NO.: 23[illegible]

VIT

Vellore Institute of Technology

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - I
WINTER SEMESTER 2024-2025

SLOT: F1

Programme Name & Branch	: B.Tech- CSE, BDS
Course Code and Course Name	: BCSE208L – Data Mining
Faculty Name(s)	: Dr. Dheeba J, Dr. Balamurugan R, Dr. Subramaniyaswamy V
Class Number(s)	: VL2024250501784, VL2024250501791, VL2024250501781
Date of Examination	: 01-Feb-25
Exam Duration	: 90 minutes

Maximum Marks: 50

General instruction(s): Answer All Questions

Q. No	Question	M	CO	BL
1.	a. Discuss the various data warehouse models and their applications in real-world. Enumerate, how the integration of these models can help in decision-making processes for large organizations. (6 Marks) b. Describe the lattice of cuboids and how it supports multidimensional data analysis. (4 Marks)	10	CO1	2
2.	A retail company's data warehouse consists of the following dimensions: Product, Store, Time, and Customer. The data in the warehouse has the year wise sales information. The company wants to analyze sales performance over time, across different stores, and for various customers. a. Draw a snowflake schema diagram for the data warehouse. b. Starting with the base cuboid, what specific OLAP operations should one perform in order to list the total sales amount per quarter for each product category.	10	CO1	3
3.	(a) Use the KDD process to design a workflow for analysing an university's student performance data. (5 Marks) (b) Consider the given data as two-dimensional data points. Given a new data point, x = (1.4,1.6) as a query, rank the database points based on similarity with the query using Euclidean distance. (5 Marks)	10	CO2	3
4.	Consider the below table of an employee record.	10	CO2	3
5.	(a) Consider the following data: [85, 23, 78, 31, 74, 45, 71, 50, 63, 59, 55, 60]. Apply equal-width binning and smooth data by bin means, boundaries and median by considering the number of bins as 4. (5 Marks) (b) Suppose two stocks A & B have the following values in a week: (2,5), (3,8), (5,10), (4,11), (6,14). Are both stocks positively or negatively correlated? (5 Marks)	10	CO2	3

	A1	A2
X1	1.5	1.7
X2	2	1.9
X3	1.6	1.8
X4	1.2	1.5
X5	1.5	1.0

EMP ID	Age	Salary	Year of Exp.
1	30	45,000	5
2	40	50,000	10
3	25	35,000	3
4	45	60,000	15
5	35	40,000	7
6	50	70,000	20

- Use min-max normalization to transform the value 35 for age onto the range [0, 1]
- Use z-score normalization to transform the value 45000 for salary.
- Use normalization by decimal scaling on the column attribute 'Year of experience'