



VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REG.NO.:

SLOT: B1

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - I
FALL SEMESTER 2025-2026

Programme Name & Branch : B.Tech CSE (Specialization in Data Science)
Course Code and Course Name : BCSE208L - Data Mining
Faculty Name(s) : Prof. Dheeba J, Prof. Gunavathi C
Class Number(s) : VL2025260101836, VL2025260101847
Date of Examination : 18.8.25
Exam Duration : 90 minutes Maximum Marks: 50

General instruction(s):

- Answer All Questions
- M - Max mark; CO - Course Outcome; BL - Blooms Taxonomy Level (1 - Remember, 2 - Understand, 3 - Apply, 4 - Analyse, 5 - Evaluate, 6 - Create)
- Course Outcomes:
CO1: Interpret the contribution of data warehousing and data mining to the decision support systems.
CO2: Construct the data needed for data mining using preprocessing techniques.

Q. No	Question	M	CO	BL
1.	A data warehouse consists of three dimensions namely Time, Doctor, and Patient. The two measures are Count and Charge, where Charge is the fee that a doctor charges a patient for a visit. (i) Draw the snowflake schema used for modeling the data warehouse. (ii) Construct the lattice of cuboids. (iii) How Roll-up and Drill-down operations could be performed on the "Time" dimension? Explain them with a sample query for each.	4 3 3	1	3
2.	Describe the three-tier architecture of a data warehouse with a neat diagram.	10	1	2
3.	How can data mining benefit a company? Provide specific examples of how techniques such as classification, clustering, association rule mining, and anomaly detection can be applied?	10	1	2
4.	a) Why is normalization needed in data preparation? b) Normalize the following values using decimal scaling. 4, -1537, 23, 0.82, -547 c) The minimum and maximum values for the attribute income are \$13,650 and \$97,300 respectively. Find the normalized value using min-max method for an income of \$55,500. The new range is [0, 1].	2 5 3	2	3
5.	Consider the data points [60, 152, NaN, 15, 53, NaN, 51, 210, 46, NaN, 41, 185]. Fill the missing values with mean value and perform the following. a) Apply equal-frequency binning to smooth the data using bin median and the bin boundaries by considering the number of bins as 4. b) Apply equal-width binning to smooth the data using bin means by considering the number of bins as 3.	5 5	2	3