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

REG.NO.: 0433

SLOT: E1+TE1

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

Programme Name & Branch : B.Tech – COMPUTER SCIENCE AND ENGINEERING
Course Code and Course Name : BCSE334L and Predictive Analytics
Faculty Name(s) : Dr. Helen Sharmila A
Class Number(s) : VL2025260502245
Date of Examination : 20.03.2026
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: (CO -2 Describe different types of predictive models and CO-3 Apply regression and classification model on applications for decision making and evaluate the performance.)

Q. No	Question	M	CO	BL
1.	A financial services company wants to build a model to predict which customers will default on their loan within the next 6 months. They have data on 50,000 customers — including demographics, spending patterns, credit score history, and past defaults. The data scientist suggests using a Propensity Model. However, the business team argues that a Clustering Model would be better. Critically evaluate both perspectives. Justify which approach is more appropriate and explain how you would validate your choice.	10	2	3
2.	Distinguish between behavioural clustering, product-based clustering, and brand-based clustering. A retail company has access to all three cluster models. Propose a framework for deciding which cluster model to apply in each of the following scenarios: (i) designing a loyalty program, (ii) planning a new product launch, (iii) reducing customer churn.	10	2	2
3.	a) A data science team compares two models: Model A has $R^2 = 0.92$ on training data and $R^2 = 0.55$ on test data. Model B has $R^2 = 0.72$ on training and $R^2 = 0.68$ on test data. The team manager says 'Model A is better because it has a higher R^2.' Critically evaluate this claim. What specific phenomenon does Model A exhibit? What metrics and diagnostic plots would you use to confirm your analysis? Which model would you deploy in production and why?	5	3	3
	b) Formally derive the decomposition of Expected MSE into three components: irreducible noise (σ^2), squared model bias, and model variance. Explain under what assumptions this decomposition holds. For a model with MSE = 25, where $\sigma^2 = 9$ and $Bias^2 = 7$, compute the model variance. What does this tell you about the model's behavior on unseen data?	5	3	3
4.	Consider three models evaluated on the same test set: Model A: MAE = 5, RMSE = 5.1 Model B: MAE = 5, RMSE = 12.6 Model C: MAE = 6, RMSE = 6.2 For a medical dosage prediction system where large errors are life-threatening, rank these models and justify each ranking decision. Which metric is more relevant here — MAE or RMSE — and why does the difference between MAE and RMSE in Model B signal a structural problem?	10	3	3

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

REG.NO.:

SLOT: E1+TE1

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

5.	(a) A model has 95% accuracy. Critically evaluate whether this is sufficient evidence to conclude that the model is performing well. Explain what sensitivity and specificity measure and describe a situation where high accuracy is misleading. (b) Why should sensitivity and specificity be reported separately instead of using a single accuracy number? Give an example were treating both errors as equally important would cause a real problem. (c) What is Youden's J Index and what does it measure? State one situation where it is useful and one situation where it is not enough on its own.	10	3	1

****