



VIT

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

REG.NO.: 23BDS0008

SLOT: B1+TB1

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - I
FALL SEMESTER 2026-2027

Programme Name & Branch : B.Tech. CSE (Data Science)
Course Code & Course Name : BCSE334L- Predictive Analytics
Faculty Name(s) : Dr ARUNKUMAR T, Dr SANTHI K, Dr SANJIBAN SEKHAR ROY
Class Number(s) : VL2026270102457, VL2026270102456, VL2026270102455
Date of Examination : 10.08.2026
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:
CO1- Understand the importance of predictive analytics and processing of data for analysis.
CO2 - Describe different types of predictive models
CO3 - Apply regression and classification model on applications for decision making and evaluate the performance.

performance.		Question	M	CO	BL																																			
Q. No																																								
1.	a) A telecom company collects customer call duration, recharge frequency, internet usage, complaint history, and payment records. Explain how predictive analytics can be used to identify customers who are likely to discontinue their subscription. Discuss the predictive model, business benefits, and decision-making process with suitable examples.	5	1	2																																				
	b) An online gaming application records players' scores, completed levels, playing time, failed attempts, in-game purchases, preferred characters and interaction patterns. The application also uses an AI-based virtual assistant that understands players' difficulties and provides personalised guidance. Using this scenario, explain how descriptive, diagnostic, predictive, prescriptive and cognitive analytics can be applied to analyse players' past performance, identify reasons for repeated failures, predict future behaviour, recommend suitable actions and provide human-like personalised assistance. For each type of analytics, state its purpose, expected output and one action the gaming company can take to improve player engagement and retention.	5																																						
2.	Consider the following dataset containing four predictor variables and one class label. <table><tr><th>Patient</th><th>Age</th><th>Gender</th><th>Blood Pressure</th><th>Cholesterol</th><th>Risk</th></tr><tr><td>1</td><td>45</td><td>1</td><td>120</td><td>180</td><td>High</td></tr><tr><td>2</td><td>52</td><td>2</td><td>135</td><td>210</td><td>Low</td></tr><tr><td>3</td><td>39</td><td>1</td><td>118</td><td>175</td><td>High</td></tr><tr><td>4</td><td>60</td><td>2</td><td>145</td><td>220</td><td>Low</td></tr><tr><td>5</td><td>50</td><td>1</td><td>130</td><td>195</td><td>High</td></tr></table> Using the Spatial Sign transformation, calculate the transformed values for all predictor variables and construct the transformed dataset. Briefly explain how the transformation reduces the effect of differences in feature magnitude.	Patient	Age	Gender	Blood Pressure	Cholesterol	Risk	1	45	1	120	180	High	2	52	2	135	210	Low	3	39	1	118	175	High	4	60	2	145	220	Low	5	50	1	130	195	High	10	2	3
Patient	Age	Gender	Blood Pressure	Cholesterol	Risk																																			
1	45	1	120	180	High																																			
2	52	2	135	210	Low																																			
3	39	1	118	175	High																																			
4	60	2	145	220	Low																																			
5	50	1	130	195	High																																			



VIT

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

REG.NO.:

SLOT: B1+TB1

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - I
FALL SEMESTER 2026-2027

3. A hospital continuously monitors patients admitted to the ICU. One patient's Blood Pressure value is missing.

Patient	Age	Heart Rate	Respiratory Rate	Oxygen Saturation (%)	Blood Pressure (mmHg)
P1	45	80	18	97	---
P2	50	82	20	96	128
P3	43	78	17	98	122
P4	55	88	22	95	136
P5	47	81	19	97	125

Using the K-Nearest Neighbours algorithm with $k = 3$, estimate the missing Blood Pressure value for Patient P1.

4. a) A music-streaming application records users' listening history, favourite artists, skipped songs, playlists, listening duration and subscription plans. Explain how a propensity model can predict whether a user is likely to listen to a newly released song or album and how these predictions can be used to personalise recommendations, improve user satisfaction and retain subscribers.

b) Calculate the Cosine Similarity and Cosine Distance for the following two documents.

Document 1

"Nanoplastics from food can enter the human bloodstream and pose health risks."

Document 2

"Nanoplastics detected in blood and brain tissues may threaten human health."

5. A hospital wants to predict post-surgical recovery time using age (X_1), surgery duration (X_2), anaesthesia duration (X_3), and mobility score (X_4).

Patient	(X_1): Age	(X_2): Surgery Duration	(X_3): Anaesthesia Duration	(X_4): Mobility Score	(Y): Recovery Time
P1	25	60	72	8	5
P2	60	90	106	4	9
P3	35	75	90	6	6
P4	50	120	140	9	14
P5	42	105	123	5	10

Using the Pearson correlation coefficient, calculate $r_{X_2 X_3}$, $r_{X_2 Y}$ and $r_{X_3 Y}$. Apply $|r| \geq 0.85$ as the threshold to identify highly correlated predictors and determine which predictor should be removed. Justify your selection, explain the problem caused by retaining both predictors, list the final predictor set, and state two benefits of removing highly correlated predictors.