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

School of Computer Science and Engineering

Fall Semester 2024-25

CAT I

SLOT: E1+TE1, E2+TE2

Programme Name & Branch: B.Tech. CSE with Specialisation in Data Science

Course Name & Code: Predictive Analytics & BCSE334L

Class Number (s): VL2024250101763, VL2024250101774, VL2024250101778, VL2024250101781, VL2024250101768

Faculty Name (s): Prof. SANTHI K, Prof. VIJAYASHERLY V, Prof. GAYATHRI P, Prof. SUBRAMANIYASWAMY V, Prof. ANBARASI M

Exam Duration: 90 Min. **Maximum Marks:** 50

General instruction(s): Answer ALL the Questions

Q.No.	Question	Max Marks
1.	Consider a super market with multiple branches in the city using software for their daily operations. Discuss any five decisions the organization can take based on the business analytics techniques applied to their customer database.	10
2.	(a) Suppose that the minimum and maximum values for the attribute income are \$11,000 and \$95,000, respectively. Find min-max normalization value for an income of \$72,500. (Note: Range is [0.0,1.0]) Suppose that the attribute income's mean and standard deviation values are \$54,000 and \$16,000, respectively. Find the z-score normalized value for the income of \$73,600. (b) In real-world data, tuples with missing values for some attributes are a common occurrence. Describe any three methods for handling this problem with examples.	10
3.	(i) Discuss any two real-time applications of propensity model which tries to predict customer churn rate. (ii) Consider an application that deals with a propensity model that tries to predict customer engagement in an E-Commerce website. List any four predictions that can be made.	10
4.	Given a utility matrix, representing the ratings, on a 1–5 star scale(with some missing values represented as NaN), of eight items, a through h, by three users A, B, and C. Compute the following from the data of this matrix. (i) Treating the utility matrix as Boolean, compute the Jaccard distance between each pair of users. (ii) Normalize the matrix by subtracting from each nonblank entry the average value for its user. Compute the cosine distance between each pair of users.	10
5.	Consider the example below where the mass, y (grams), of a chemical is related to the time, x (seconds), for which the chemical reaction has been taking place according to the table: Find the equation of the regression line. What is the mass of the chemical after ten seconds has passed? By how much does the chemical increase in weight in five seconds?	10

	a	b	c	d	e	f	g	h
A	4	5		5	1		3	2
B		3	4	3	1	2	1	
C	2		1	3		4	5	3

Time, x (seconds)	5	7	12	16	20
Mass, y (grams)	40	120	180	210	240