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

REG.NO.: [redacted]

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

SLOT: C2+TC2

Programme Name & Branch	: M.Tech. (CSE)- 5 Year Integrated
Course Code and Course Name	: Advanced Predictive Analytics & MDI3003
Faculty Name(s)	: Dr. ARCHANA T, Dr. BALAJI G N, Dr. UMA PRIYA D
Class Number(s)	: VL2024250502378/2376/2380
Date of Examination	: 29.01.2025
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:
  1. Understand the process of formulating objectives, data selection/collection, preparation and process to successfully design the model.
  2. Able to prepare and process data for the models.
  3. Gain the insights from the data through Exploratory Data Analysis for feature engineering.

Q. No	Question	M	CO	BL
1.	a) Is Predictive Analytics the same as Statistics? Justify your reasons.	5	CO1	5
	b) What challenges are commonly faced in data modeling? Explain.	5		2
2.	In the CRISP-DM process, data understanding and data preparation are critical phases. Explain the main tasks involved in these phases and explain them for the use case **Predicting Loan Defaults for a Bank**	10	CO1	1
3.	You are provided with a dataset containing information about individuals, including missing values, incorrect data, and inconsistent formatting. The dataset is as follows:	10	CO2	4

ID	Name	Email	Age	Ph. Number
1	John Doe	john.doe@gmail.	25	123-456-7890
2	Jane Smith	N/A	NaN	(555)-123-4567
3	Mike Taylor	mike.taylor@website	-30	555 555 555
4	Ann Brown	annbrown@@yahoo.com	29	N/A
5	Chris Green	chris.green@gmail.com	45	9876543210

Perform the following data-cleaning tasks: i) Handle Missing Values
ii) Rectify Incorrect Data iii) Standardize the data
Provide the final cleaned dataset and explain the steps you took to clean it.

4.	Explain the steps involved in building a predictive analytics model, starting from data preprocessing to model validation for the use case **Predicting Customer Churn for a Subscription-Based Streaming Service**	10	CO2	3

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

REG.NO.:

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

SLOT: C2+TC2

5. Consider the following dataset for the case study **Employee Performance and Compensation Analysis.** Draw Scatter Plots and Bar Plots wherever necessary. (Bivariate Analysis) — 10 | CO3 | 3

Name	Department	Exp. (Years)	Education	Perf.Rating	Salary ($)
John	IT	5	Graduate	8	75,000
Alice	Sales	3	Non-Graduate	7	55,000
Bob	Finance	7	Graduate	9	85,000
Eve	HR	4	Graduate	6	45,000
Charlie	IT	6	Non-Graduate	9	80,000

a. How does an employee's years of experience influence their salary? Use scatter plots to visualize the insights.
b. Is there a relationship between employee performance ratings and their salaries? Analyze this trend and identify any deviations or outliers from the pattern using scatter plots.
c. How is educational qualification distributed across different departments? Are there any noticeable patterns in the qualifications of employees in various departments? Visualize this using bar plots.
d. Which departments have the highest average salaries? How does the department of an employee influence their salary? Analyze and visualize the insights using bar plots.

****