REG.NO.:

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

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

SLOT: C2 + TC2

Programme Name & Branch	**: M.Tech (Integrated) & Computer Science and Engineering**
Course Code and Course Name	**: MDI3003 and Advanced Predictive Analytics**
Faculty Name(s)	**: Dr. G. N. Balaji, Dr. Archana T, Dr. UMA PRIYA D**
Class Number(s)	**: VL2024250502376, 2378, 2380**
Date of Examination	**: 18.03.2025**
Exam Duration	**: 90 minutes**

Maximum Marks: 50

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)
- CO3: Gain the insights from data through Exploratory data analysis for feature engineering
- CO4: Compare the underlying predictive modeling techniques. Analyse the performance of the model and quality of results.
- CO5: Explore predictive models to enhance the prediction performance.

<table>
<tr><th>Q. No</th><th>Question</th><th>M</th><th>CO</th><th>BL</th></tr>
<tr><td>1.</td><td>Given data = {2.5, 0.5, 2.2, 1.9, 3.1, 2.3, 2.0, 1.0, 2.4, 0.7; 2.9, 2.2, 3.0, 2.7, 1.6, 1.1, 1.6, 0.9}. apply Principal Component Analysis (PCA) to transform the data into a new coordinate system where the most significant variations are captured.</td><td>10</td><td>3</td><td>3</td></tr>
<tr><td>2.</td><td>A family enjoys different weekend activities such as going to the cinema, playing tennis, shopping, or staying in. Their choice depends on several factors: the weather, whether parents are present, and their financial status. However, they do not follow a strict routine, making it challenging to predict their decisions. Use the given data, to uncover the decision-making pattern of the family. By applying a decision tree model, determine how different factors influence their weekend plans.
<table>
<tr><th>Weekend</th><th>Weather</th><th>Parents</th><th>Money</th><th>Decision</th></tr>
<tr><td>W1</td><td>Sunny</td><td>Yes</td><td>Rich</td><td>Cinema</td></tr>
<tr><td>W2</td><td>Sunny</td><td>No</td><td>Rich</td><td>Tennis</td></tr>
<tr><td>W3</td><td>Windy</td><td>Yes</td><td>Rich</td><td>Cinema</td></tr>
<tr><td>W4</td><td>Rainy</td><td>Yes</td><td>Poor</td><td>Cinema</td></tr>
<tr><td>W5</td><td>Rainy</td><td>No</td><td>Rich</td><td>Stay In</td></tr>
</table></td><td>10</td><td>4</td><td>5</td></tr>
<tr><td>3.</td><td>A bank wants to automate its loan approval process based on applicants' profiles. The bank considers three categorical factors: Employment Status (Employed/Unemployed), Credit Score (Good/Poor), and Existing Debt (Yes/No). Using historical loan data, the bank aims to predict whether a new applicant's loan should be approved or rejected. Assume K = 3</td><td>10</td><td>4</td><td>3</td></tr>
</table>

Dataset:

Applicant	Employment	Credit Score	Existing Debt	Loan Status
A1	Employed	Good	No	Approved
A2	Unemployed	Good	No	Approved
A3	Employed	Poor	No	Approved
A4	Employed	Good	Yes	Approved
A5	Unemployed	Poor	No	Rejected
A6	Employed	Poor	Yes	Rejected
A7	Unemployed	Poor	Yes	Rejected
A8	Employed	Good	No	Approved
A9	Unemployed	Good	Yes	Approved
A10	Employed	Good	No	Approved

4.	In a financial fraud detection system, false negatives (fraudulent transactions classified as legitimate) are more costly than false positives (legitimate transactions flagged as fraud). Given this scenario, how would you design an ensemble model using bagging and boosting to optimize detection performance? Explain how each technique contributes to reducing misclassification and which method you would prioritize for this application.	10	5	2
5.	A healthcare company is developing an AI-based diagnostic system to detect diabetes based on patient data. The dataset contains features such as Glucose Level, BMI and Age. To improve prediction accuracy, consider using a heterogeneous ensemble model that combines Naïve bayes and K-NN.	10	5	6

Patient	Glucose Level	BMI	Age	Diabetes (Target)
P1	180	30	45	1
P2	90	22	35	0
P3	150	28	50	1
P4	85	24	40	0
P5	200	33	55	1

classify a new patient (P6) with: Glucose Level = 160, BMI = 27, Age = 48

****