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

REG.NO.:

SLOT: D2

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

Programme Name & Branch	**: B.Tech & CSE and BS**
Course Code and Course Name	**: CBS3006, Machine Learning**
Faculty Name(s)	**: Dr. A. Karthikeyan, Dr. Nageshwar Rao R**
Class Number(s)	**: VL2025260502105, VL2025260502106**
Date of Examination	**: 18.03.26**
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: (Type the CO statements covered in this question paper. Use the CO number as per the syllabus copy)
  CO3 Solve real world applications by selecting suitable learning model.
  CO4 Boost the performance of the model by combining results from different approaches.
  CO5 Recognize and classify sequencing patterns using HMM.

<table>
<tr><th>Q. No</th><th>Question</th><th>M</th><th>CO</th><th>BL</th></tr>
<tr><td>1.</td><td>a) Analyse the dataset given below and identify the top three players required to win a match. Use the K-Nearest Neighbors Algorithm. Two different sets of test data are used for the same analysis. The test data (i) Matches played =945, Goals scored=625, (ii) Matches played =865, Goals saved=532. Show the results in a tabular form with Player name, Position, and Country.
<table>
<tr><th>Players</th><th>Matches played</th><th>Goals scored</th><th>Goals saved</th><th>Position</th><th>Country</th></tr>
<tr><td>Cristiano</td><td>1300</td><td>950</td><td>395</td><td>Forward</td><td>Portugal</td></tr>
<tr><td>Messi</td><td>1130</td><td>890</td><td>405</td><td>Forward</td><td>Argentina</td></tr>
<tr><td>Ronaldo</td><td>616</td><td>513</td><td>286</td><td>Forward</td><td>Brazil</td></tr>
<tr><td>Pirlo</td><td>1008</td><td>700</td><td>653</td><td>Midfielder</td><td>Italy</td></tr>
<tr><td>Ramos</td><td>1002</td><td>256</td><td>873</td><td>Defender</td><td>Spain</td></tr>
<tr><td>Lucas</td><td>986</td><td>231</td><td>745</td><td>Defender</td><td>France</td></tr>
</table></td><td>6</td><td>3</td><td>4</td></tr>
<tr><td></td><td>b) Consider an event between two teams, say team A and team B. Suppose team A wins 56% of the time and team B wins the remaining events. Among the events won by team A, only 32% of the events had occurred at team B's location. At the same time, 68% of the victories for team B are obtained at their own location. If team B is hosting the next event between the two teams, what is the probability that team B emerges as the winner? Also, identify the winner.</td><td>4</td><td></td><td></td></tr>
<tr><td>2.</td><td>Draw an optimal hyperplane for the linearly separable data points given below. The points (4, 1) (4, -1) and (6, 0) belong to the positive class, and the points (1, 0) (0, 1) and (0, -1) belong to the negative class.</td><td>10</td><td>3</td><td>3</td></tr>
</table>

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

REG.NO.:

SLOT: D2

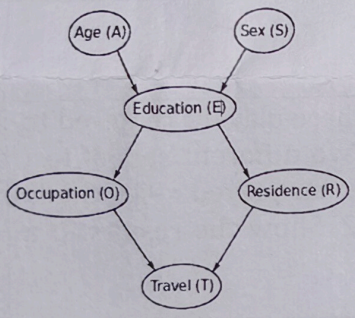
SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

---

3. Find the optimal decision tree using the CART algorithm and hence predict the possible decision for the dataset given below. — **10** — CO: 3 — BL: 3

City size	Avg. Income	Investors	Awareness	Decision
Big	High	Yes	High	Yes
Medium	Medium	No	Medium	No
Small	Low	Yes	Low	No
Big	High	No	High	Yes
Small	Medium	Yes	High	No
Medium	High	Yes	Medium	Yes
Medium	Medium	Yes	Medium	No
Big	Medium	No	Medium	No
Medium	High	Yes	Low	No
Small	High	No	High	Yes
Small	Medium	No	High	No
Medium	High	No	Medium	No

4. a) Calculate all the conditional probabilities of travel for the given diagram using the Bayesian model. The observed probabilities are P(Age(A)) = (Young=0.30, Adult=0.50, old=0.20) and P(Sex(s)) = (M=0.60, F=0.40) — **6** — CO: 4 — BL: 3



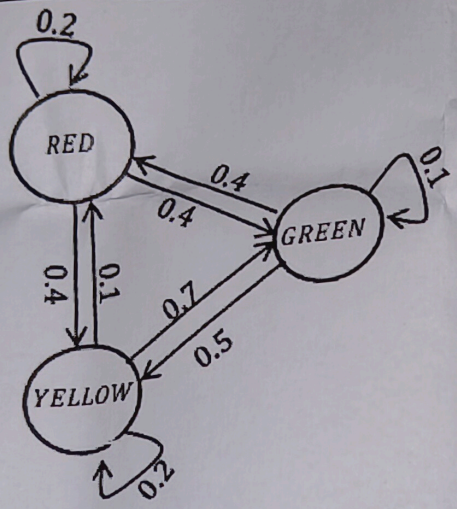
b) A scan report used to predict fever is stable (low variance), and it does not show the critical nature of the disease and leads to more false rejection rate. Error is high during training and testing. Which boosting or Bagging technique would you recommend for this scenario? Why? Justify the same. — **4**

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

REG.NO.:

SLOT: D2

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

5.				10	5	4

a) Using the sample probabilities in the given figure, start distribution π = [0.3,0.4,0.3], compute the probability of the sequence: [YELLOW, GREEN, RED] **[4 Marks]**

b) Given the emission probabilities for the observations [DARK, LIGHT], compute the probability of the observation sequence O = [DARK, LIGHT], *using forward computation.* **[6 Marks]**

	DARK	LIGHT
RED	0.8	0.2
GREEN	0.6	0.4
YELLOW	0.4	0.6