

**School of Computer Science Engineering and Information Systems**

**Winter Semester 2024-2025**

**Continuous Assessment Test – II**

**Programme Name & Branch: MCA**

**Course Name & code: PMCA507L**

**Class Number (s): 3327 and 3331**

**Faculty Name (s): Dr. Parimala M & Dr. Anitha A**

**Exam Duration: 90 Min.**

**Maximum Marks: 50**

**General instruction(s): Answer All Questions**

**CO – Course Outcome; BL – Blooms Taxonomy Level (1 – Remember, 2 – Understand, 3 – Apply, 4 – Analyze, 5 – Evaluate, 6 – Create)**

**Course Outcomes:**

**CO2: Provide solutions for classification, regression, and clustering approaches in real world applications**

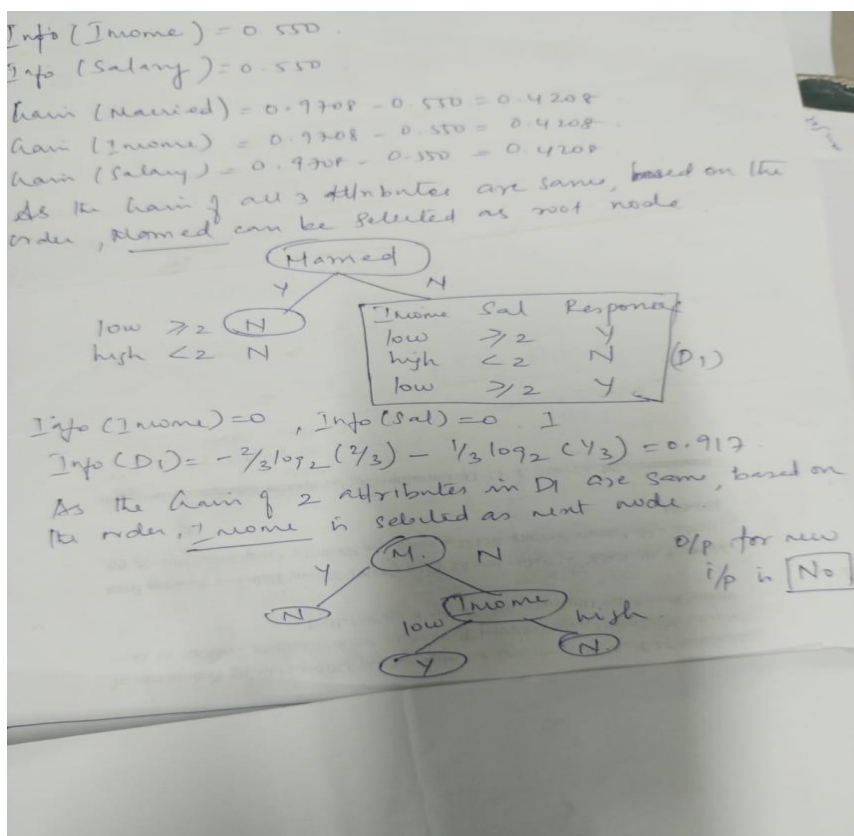
**CO3: Gain knowledge to combine machine learning models to achieve better results**

Q.No.	Question	Max Marks	CO	BL																					
1.	<p>Predict the weight of student for the dataset with five samples as given below using kNN Algorithm where the value of k is 5.</p> <table><tr><th>Age</th><th>Height</th><th>Weight (target class)</th></tr><tr><td>25</td><td>160</td><td>80</td></tr><tr><td>32</td><td>145</td><td>70</td></tr><tr><td>24</td><td>160</td><td>75</td></tr><tr><td>44</td><td>165</td><td>60</td></tr><tr><td>21</td><td>152</td><td>62</td></tr><tr><td><b>20</b></td><td><b>155</b></td><td><b>??</b></td></tr></table>	Age	Height	Weight (target class)	25	160	80	32	145	70	24	160	75	44	165	60	21	152	62	<b>20</b>	<b>155</b>	<b>??</b>	10	CO 2	BL
Age	Height	Weight (target class)																							
25	160	80																							
32	145	70																							
24	160	75																							
44	165	60																							
21	152	62																							
<b>20</b>	<b>155</b>	<b>??</b>																							

	<p>① <u>KNN</u></p> <table><thead><tr><th></th><th>Age</th><th>Height</th><th>(target) Weight</th><th>Distance</th><th>Rank</th></tr></thead><tbody><tr><td>O1</td><td>25</td><td>160</td><td>80</td><td><math>\sqrt{(25-20)^2 + (160-155)^2} = 7.07</math></td><td>2</td></tr><tr><td>O2</td><td>32</td><td>145</td><td>70</td><td><math>\sqrt{(32-20)^2 + (145-155)^2} = 15.62</math></td><td>4</td></tr><tr><td>O3</td><td>24</td><td>160</td><td>75</td><td><math>\sqrt{(24-20)^2 + (160-155)^2} = 6.40</math></td><td>3</td></tr><tr><td>O4</td><td>44</td><td>165</td><td>60</td><td><math>\sqrt{(44-20)^2 + (165-155)^2} = 26</math></td><td>5</td></tr><tr><td>O5</td><td>21</td><td>182</td><td>62</td><td><math>\sqrt{(21-20)^2 + (182-155)^2} = 3.16</math></td><td>1</td></tr></tbody></table> <p>For new i/p. Age = 20, height = 152, weight = ??</p> <div>weight = 62</div>		Age	Height	(target) Weight	Distance	Rank	O1	25	160	80	$\sqrt{(25-20)^2 + (160-155)^2} = 7.07$	2	O2	32	145	70	$\sqrt{(32-20)^2 + (145-155)^2} = 15.62$	4	O3	24	160	75	$\sqrt{(24-20)^2 + (160-155)^2} = 6.40$	3	O4	44	165	60	$\sqrt{(44-20)^2 + (165-155)^2} = 26$	5	O5	21	182	62	$\sqrt{(21-20)^2 + (182-155)^2} = 3.16$	1			
	Age	Height	(target) Weight	Distance	Rank																																			
O1	25	160	80	$\sqrt{(25-20)^2 + (160-155)^2} = 7.07$	2																																			
O2	32	145	70	$\sqrt{(32-20)^2 + (145-155)^2} = 15.62$	4																																			
O3	24	160	75	$\sqrt{(24-20)^2 + (160-155)^2} = 6.40$	3																																			
O4	44	165	60	$\sqrt{(44-20)^2 + (165-155)^2} = 26$	5																																			
O5	21	182	62	$\sqrt{(21-20)^2 + (182-155)^2} = 3.16$	1																																			
2.	<p>A set of training data is given in table below.</p> <table><thead><tr><th>Married</th><th>Income (K)</th><th>Salary (lakhs)</th><th>Response</th></tr></thead><tbody><tr><td>Yes</td><td>low</td><td><math>\geq 2</math></td><td>No</td></tr><tr><td>No</td><td>low</td><td><math>\geq 2</math></td><td>Yes</td></tr><tr><td>No</td><td>high</td><td><math>&lt; 2</math></td><td>No</td></tr><tr><td>Yes</td><td>high</td><td><math>&lt; 2</math></td><td>No</td></tr><tr><td>No</td><td>low</td><td><math>\geq 2</math></td><td>Yes</td></tr></tbody></table> <p>(i) Build a decision tree for the above table using Information gain technique. Assume attribute “Response” as target class.</p> <p>(ii) Predict the value of target class for the new sample “Married=Yes, Income=low, Salary=<math>\geq 2</math>” using the tree constructed in the previous step?</p>	Married	Income (K)	Salary (lakhs)	Response	Yes	low	$\geq 2$	No	No	low	$\geq 2$	Yes	No	high	$< 2$	No	Yes	high	$< 2$	No	No	low	$\geq 2$	Yes	10	CO 2	B L												
Married	Income (K)	Salary (lakhs)	Response																																					
Yes	low	$\geq 2$	No																																					
No	low	$\geq 2$	Yes																																					
No	high	$< 2$	No																																					
Yes	high	$< 2$	No																																					
No	low	$\geq 2$	Yes																																					

$$\begin{aligned}
 (2) \text{Info}(D) &= I(2,3) = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \\
 &= \boxed{0.9708} \\
 \text{Info}(\text{Married}) &= \frac{2}{5} I(0,2) + \frac{3}{5} I(1,2) \\
 &= \frac{2}{5} \left[ -\frac{0}{2} \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) \right] + \\
 &\quad \frac{3}{5} \left[ -\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) \right] = \boxed{0.550}
 \end{aligned}$$

	$c_1(4)$	$c_2(N)$	$I(c_1, c_2)$
Y	0	2	$I(0,2)$
N	1	2	$I(1,2)$



3.

SNo	Past Trend	Open interest	Trading volume	Return (Target Class)
1	Positive	Low	High	Up
2	Negative	High	Low	Down
3	Positive	Low	High	Up

10

CO  
2

B  
L

4	Positive	High	High	Up
5	Negative	Low	High	Down
6	Positive	Low	Low	Down
7	Negative	High	High	Down
8	Negative	Low	High	Down
9	Positive	Low	Low	Down
10	Positive	High	High	Up

Build 2 Decision trees using the Gini index and classify the features using random forest algorithm. Use appropriate ensemble technique to derive the final output for the input features "Past trend= positive, Open interest=High, Trading volume=low, Return=??"

Bootstrap sample 1: {1,2,3,4,5}

Bootstrap sample 2: {6,7,8,9,10}

③  $\frac{+ve(3)}{(1,2,3,4,5) \text{ PT}}$

$$Gini(PT) = 1 - \left[ \frac{3}{5} \right] = 0$$

$-ve(2)$

$$Gini(PT) = 1 - \left[ \frac{2}{5} \right] = 0$$

$$WA(PT) = \boxed{0} \checkmark$$

Open Interest

Low = 3

$$Gini(OI=low) = 1 - \left[ \left( \frac{2}{3} \right)^2 + \left( \frac{1}{3} \right)^2 \right] = 0.44$$

high = 2

$$Gini(OI=high) = 1 - \left[ \left( \frac{1}{2} \right)^2 + \left( \frac{1}{2} \right)^2 \right] = 0.5$$

$$WA = \frac{3}{5} \times 0.44 + \frac{2}{5} \times 0.5 = \boxed{0.464}$$

Trading volume

high = 4

$$Gini(TV=high) = 1 - \left[ \left( \frac{3}{4} \right)^2 + \left( \frac{1}{4} \right)^2 \right] = 0.375$$

low = 1

$$Gini(TV=low) = 1 - \left[ \left( \frac{1}{1} \right)^2 \right] = 0$$

$$WA = \frac{4}{5} \times 0.375 + \frac{1}{5} \times 0 = \boxed{0.3}$$

As the weighted average (WA) is low for PT.

```

graph TD
    PT[PT] -- +ve --> UP((UP.))
    PT -- -ve --> DOWN((down.))
  
```

$D_2 (6, 7, 8, 9, 10)$

$$Q_{min}(PT = +ve) = 1 - \left[ \left( \frac{2}{3} \right)^2 + \left( \frac{1}{3} \right)^2 \right] = 0.44$$

$$Q_{min}(PT = -ve) = 1 - \left[ \frac{2}{2} \right] = 0$$

$$WA(PT) = \frac{3}{5} \times 0.44 = \boxed{0.264}$$

$$Q_{min}(OI = low) = 1 - \left[ \frac{3}{3} \right] = 0$$

$$Q_{min}(OI = high) = 1 - \left[ \left( \frac{1}{2} \right)^2 + \left( \frac{1}{2} \right)^2 \right] = 0.5$$

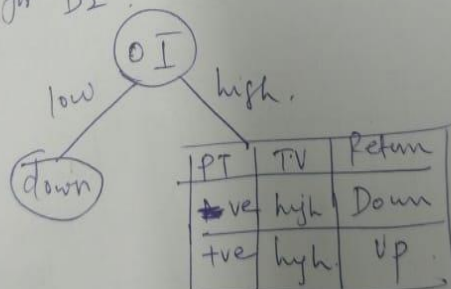
$$WA(OI) = \frac{2}{5} \times 0.5 = \boxed{0.2} \quad \checkmark$$

$$Q_{min}(TV = high) = 1 - \left[ \left( \frac{2}{3} \right)^2 + \left( \frac{1}{3} \right)^2 \right] = 0.44$$

$$Q_{min}(TV = low) = 1 - \left[ \left( \frac{2}{2} \right)^2 \right] = 0$$

$$WA(TV) = \frac{3}{5} \times 0.44 = \boxed{0.264}$$

Weighted Average of ~~the~~ open Interest is low, so OI is the root node for D2.



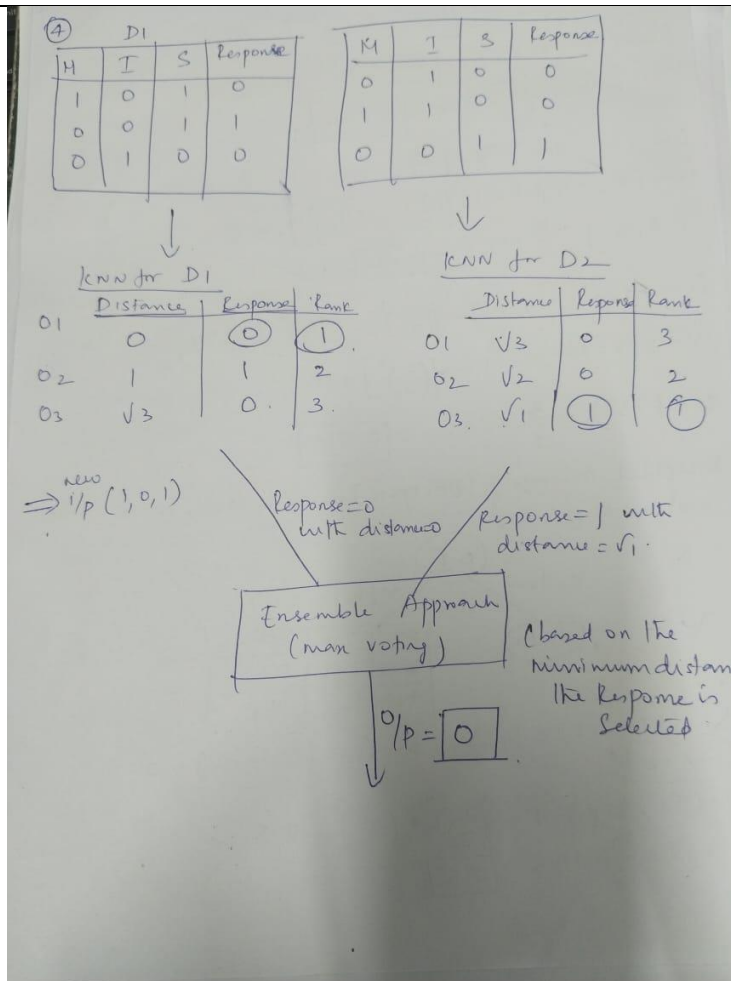
$$Q_{min}(PT = -ve) = 1 - \left( \frac{1}{1} \right) = 0. \quad Q_{min}(PT = +ve) = 0$$

$$\boxed{WA = 0}$$

$$Q_{min}(TV = high) = 1 - \left[ \left( \frac{1}{2} \right)^2 + \left( \frac{1}{2} \right)^2 \right] = 0.5$$

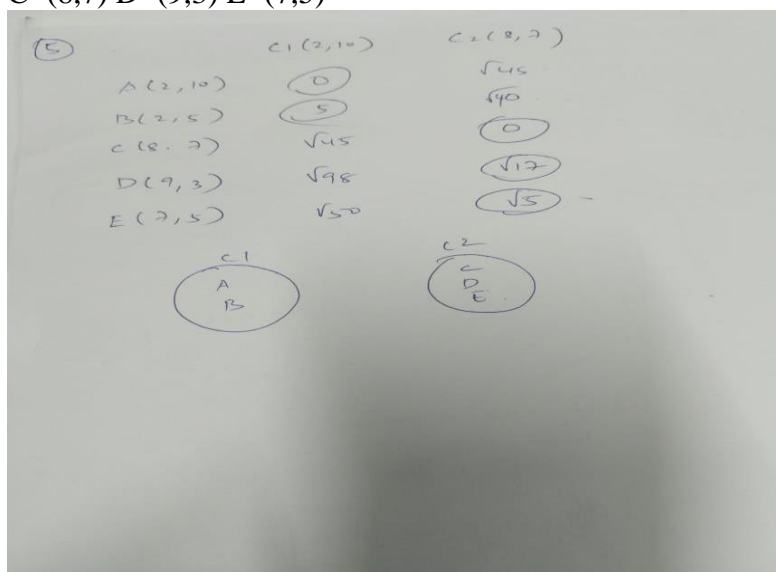
$$WA = \frac{2}{2} \times 0.5 = \boxed{0.5}$$

	<p>Random Forest - 1</p> <p>Random forest - 2</p> <p>⇒ new /p PT = true, OI = high, TV = low, Return = ?</p> <p>Response = up</p> <p>Response = up</p> <p>Ensemble approach max voting</p> <p>O/p: Response = <span style="border: 1px solid black; padding: 2px;">Up</span></p>			
4.	<p>Apply the bagging technique using kNN classifier model to the bootstrap samples and identify the class label for the input feature “Married=Yes, Income=low, Salary&gt;=2”. Consider the dataset given in Q.No.2</p> <p>Bootstrap sample 1: {1,2,3}</p> <p>Bootstrap sample 2: {3,4,5}</p>	10	CO 3	B L



5.

Group the following points using k-means algorithm into two clusters using Euclidean distance. Assume 'A' and 'C' as initial centroid. Find the clusters after one iteration. A=(2,10) B=(2,5) C=(8,7) D=(9,3) E=(7,5)



10

CO  
3

B  
L

