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

Slot: B2

School of Computer Science Engineering and Information Systems

Winter Semester 2024-2025

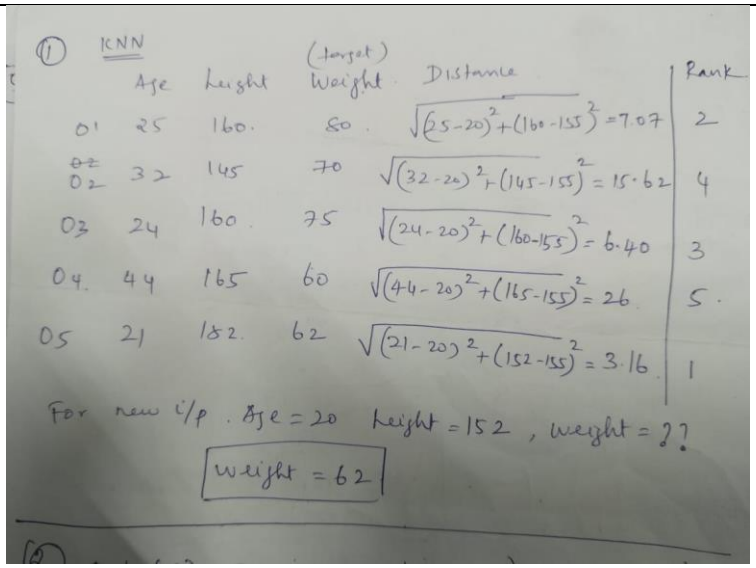
Continuous Assessment Test – II

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**

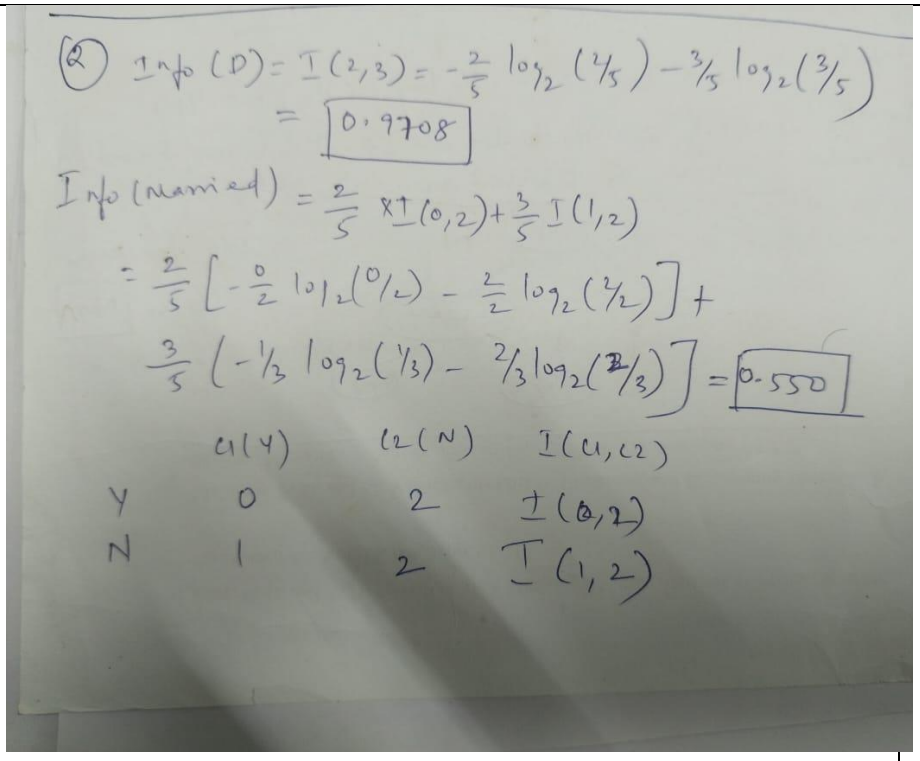
CO – Course Outcome; BL – Blooms Taxonomy Level (1 – Remember, 2 –Understand, 3 – Apply, 4 – Analyze, 5 – Evaluate, 6 – Create)
Course Outcomes:
CO2: Provide solutions for classification, regression, and clustering approaches in real world applications
CO3: Gain knowledge to combine machine learning models to achieve better results

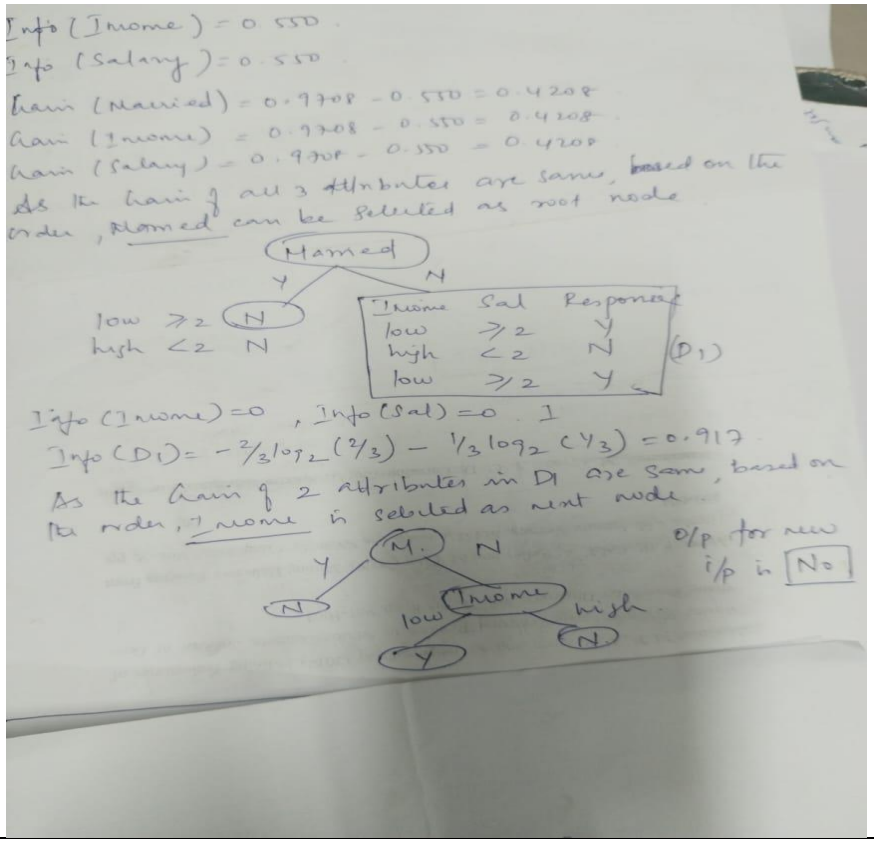
<table>
<tr><th>Q.No.</th><th>Question</th><th>Max Marks</th><th>CO</th><th>BL</th></tr>
<tr><td>1.</td><td>Predict the weight of student for the dataset with five samples as given below using kNN Algorithm where the value of k is 5.
<table>
<tr><th>Age</th><th>Height</th><th>Weight (target class)</th></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>20</td><td>155</td><td>??</td></tr>
</table>
</td><td>10</td><td>CO 2</td><td>BL</td></tr>
</table>



2.	A set of training data is given in table below. (i) Build a decision tree for the above table using Information gain technique. Assume attribute "Response" as target class. (ii) Predict the value of target class for the new sample "Married=Yes, Income=low, Salary=>=2" using the tree constructed in the previous step?	10	CO 2	B L

Married	Income (K)	Salary (lakhs)	Response
Yes	low	>=2	No
No	low	>=2	Yes
No	high	<2	No
Yes	high	<2	No
No	low	>=2	Yes



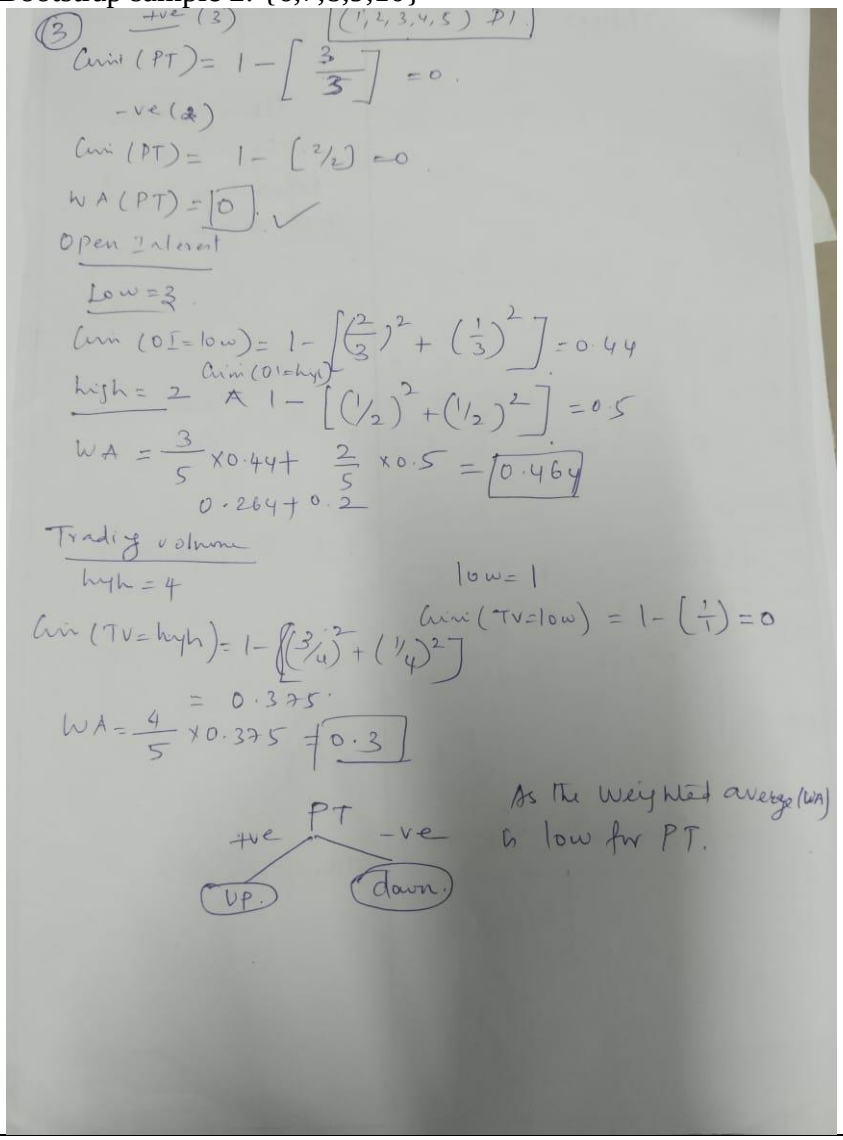


3.						10	CO 2	B L

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

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}



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

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

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

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

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

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

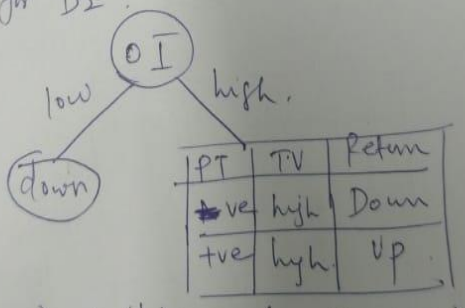
WA (OI) = $\frac{2}{5} \times 0.5 = \boxed{0.2}$ ✓

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

Gini (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 open Interest is low, so OI is the root node for D2.



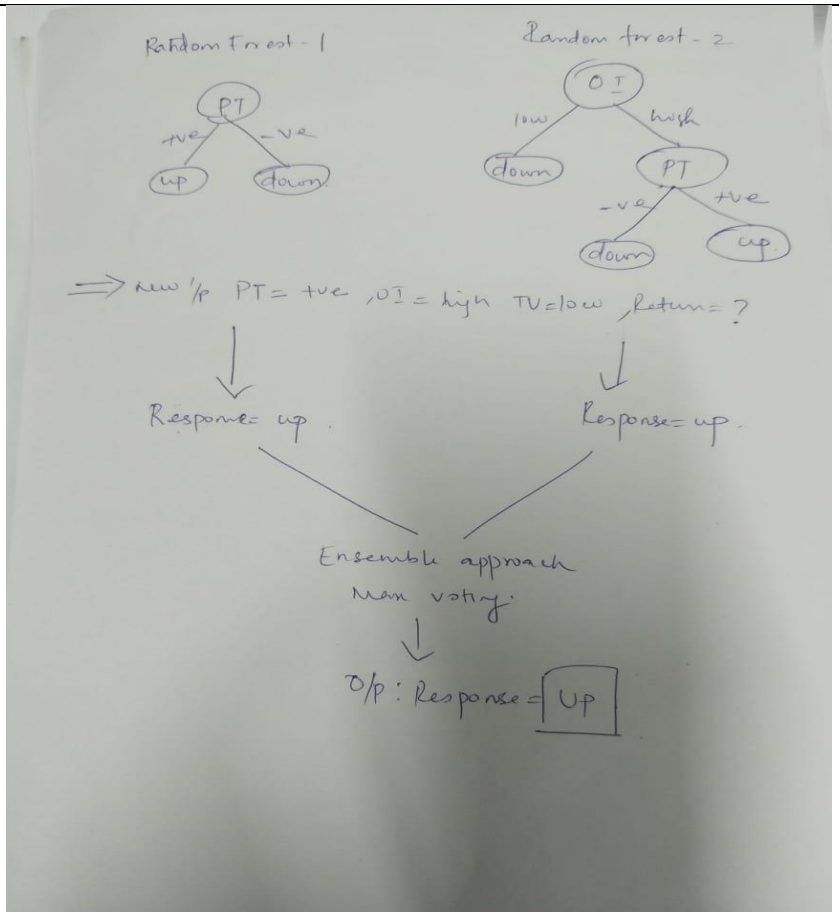
PT	TV	Return
-ve	high	Down
+ve	high	Up

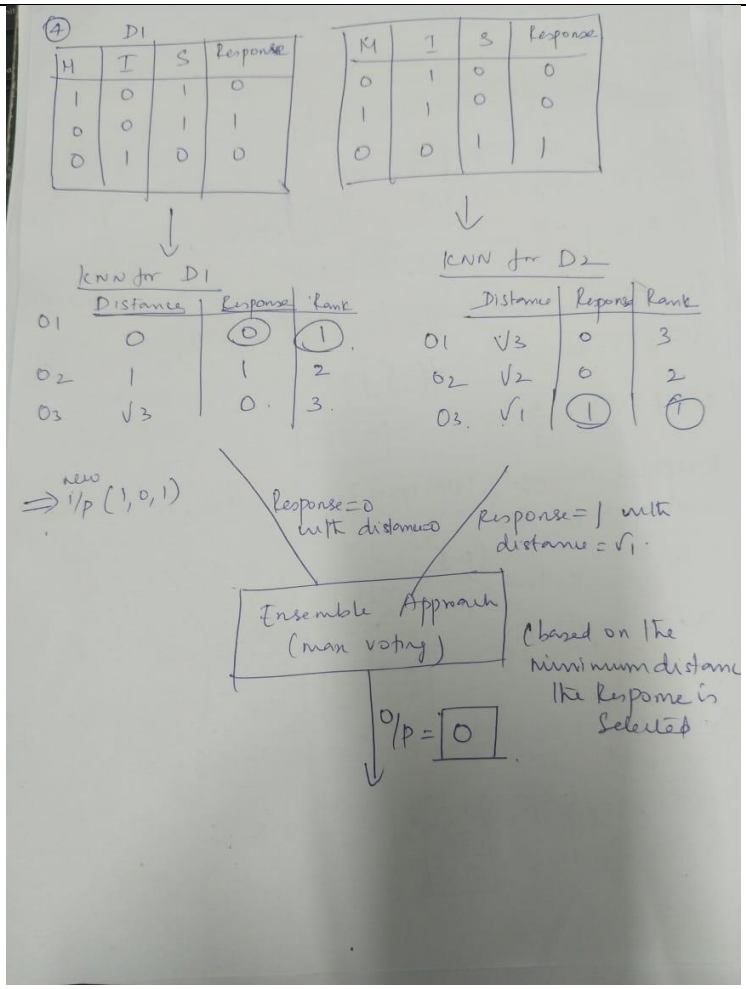
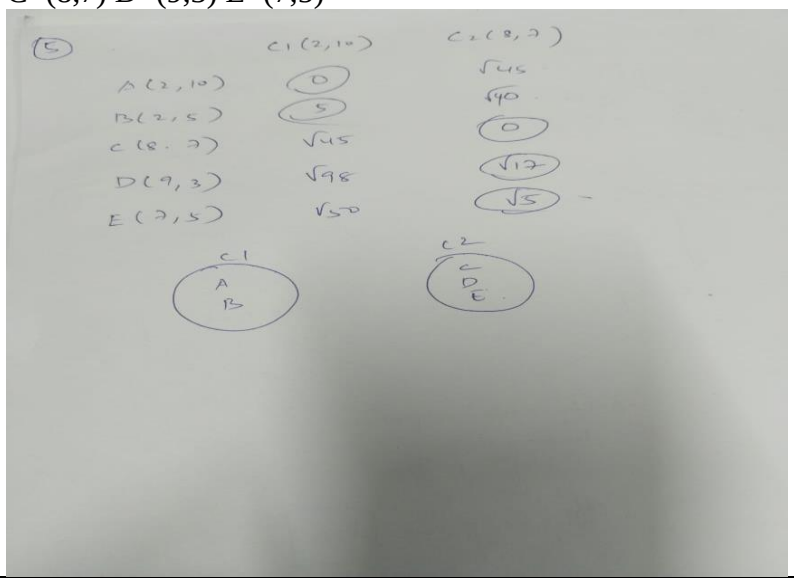
Gini (PT = -ve) = $1 - \left(\frac{1}{1}\right) = 0$. Gini (PT = +ve) = 0

$\boxed{WA = 0}$

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

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

				
4.	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=>=2" . Consider the dataset given in Q.No.2 Bootstrap sample1: {1,2,3} Bootstrap sample 2: {3,4,5}	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	BL