Question #02.

As following the order for spliting Age and cheese Content, there is the below 1st split.

$$Age = \frac{Max + Min}{2} = \frac{54+8}{2} \Rightarrow \boxed{Age = 31}$$

- Then; split on the left subtree of cheese content.

$$\text{Cheese Content} = \frac{Max + Min}{2} = \frac{9.2+6.2}{2} \Rightarrow \boxed{\text{cheese Content} = 7.7}$$

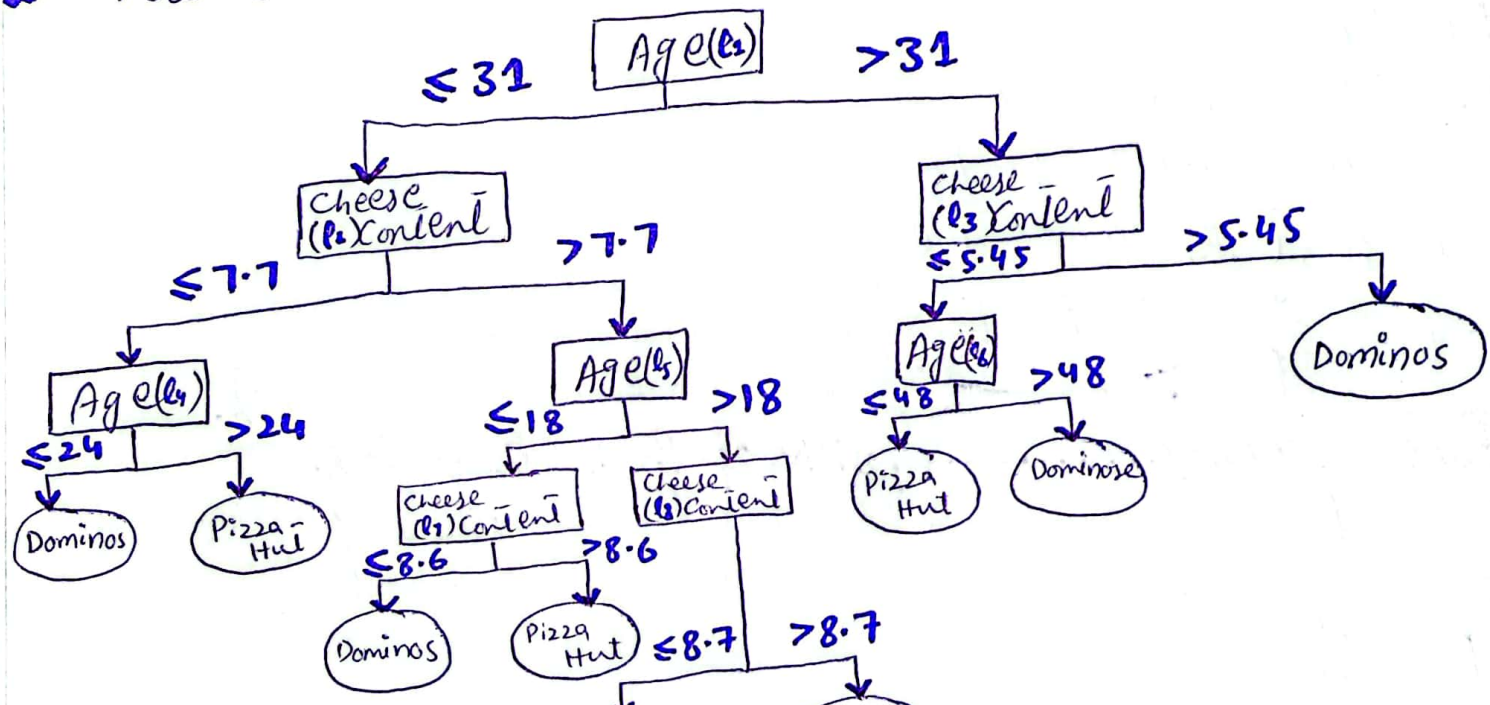
- Now; split on the right subtree of cheese content.

$$\text{Cheese Content} = \frac{Max + Min}{2} = \frac{7.6+3.3}{2} \Rightarrow \boxed{\text{Cheese Content} = 5.45}$$

- Now; Split on left subtree after cheese content.

$$Age = \frac{Max + Min}{2} = \frac{27+20}{2} \Rightarrow Age = 23.5 \cong 24 \Rightarrow \boxed{Age = 24}$$

- So; here is below K-D Tree.



So, Test data.

Name	Age	cheeseContent	Pizza Outlet
Harry	46	7	???

So; $l_1 \rightarrow l_3 \rightarrow$ Dominos

So; Results are following

Name	Age	Cheese Content	Pizza Outlet
Harry	46	7	Dominos

Question #03

a). Rule accuracy ..... none of R_1 is discarded.

if the examples for R_1 are not discarded then R_2 will be choosen because, it has higher accuracy then R_3.

Extraction from the figure are given below;

So; All positive examples = 29

All negative examples = 21.

R_1 has 12 positive, 3 negative.

R_2 has 7 positive, 3 negative.

R_3 has 8 positive, 4 negative.

$$\text{Accuracy} = \frac{\text{Positive results}}{\text{Total results}}$$

$R_1 = 12/16 = 0.8 = 80\%$

$R_2 = 7/10 = 0.7 = 70\%$

$R_3 = 8/12 = 0.667 = 66.7\%$

R_1 is best, than R_2 and R_3.

So; if we discard R_1 then R_2 is best option.

$R_2 > R_3$ (because R_2 has higher accuracy than R_3).

b). Rule accuracy ..... positive of R_1 is discarded.

if R_1 positive examples are discarded, then accuracy of R_2 and R_3 is affected because R_1 and R_3 are overlapped.

$R_2 = 7/10 = 0.7 = 70\%$.

$R_3 = 6/10 = 0.6 = 60\%$.

So; in this case there is also better to choose R_2 in this case.

$R_2 > R_3$. (because R_2 has more accuracy than R_3)

d). Rule accuracy....., when both positive and negative discard.

If the both positive and negative examples of R_1 is discarded.

So; accuracy updated:

$R_2 = 7/10 = 0.7 = 70\%$ $R_3 = 6/8 = 0.75 = 75\%$.

So; in this case we preferred R_3 over R_2. (because R_3 has more accuracy than R_2).

$R_3 > R_2$

Question #04.

Step 1: Calculate distance. (L_2 Norm is Euclidian Distance).

$d_6 = \sqrt{(120000 - 90000)^2} \Rightarrow \boxed{d_6 = 30000}$ ✓

Question #05.

Confusion Matrix

a). Total Number Of Instances.

Sum Of All Values

Roses	100	10	5
Daises	15	85	20
Tulips	8	18	90
Total	123	113	115

Total = 123 + 113 + 115

Total = 351

b). Accuracy Computation.

$$\text{Accuracy} = \frac{\text{Total Correct predictions}}{\text{Total No. of predictions}} = \frac{TP + TN}{\text{Total instances}} = \frac{100 + 85 + 90}{351} = \frac{275}{351}$$

Accuracy = 0.7834 $\Rightarrow$ **Accuracy = 78.34 %**

c). Sensitivity (Recall).

$$\frac{TP}{TP + FN} \Rightarrow \text{Roses} = \frac{100}{100 + 10 + 5} = \frac{100}{115} = 0.869 \Rightarrow \textbf{Roses = 86.9 \%}$$

$$\text{Daises} = \frac{85}{15 + 85 + 20} = \frac{85}{120} = 0.7083 \Rightarrow \textbf{Daises = 70.83 \%}$$

$$\text{Tulips} = \frac{90}{8 + 18 + 90} = \frac{90}{116} = 0.7758 \Rightarrow \textbf{Tulips = 77.58 \%}$$

Roses highest sensitivity indicates that best performance is also be "correctly identifying" instance of each class.

Daises has lower sensitivity which indicates that is "incorrectly identified"

d) Specificity

$$TNR = \frac{TN}{TN + FP}$$

$$Spec_{Setosa} = \frac{85 + 20 + 12 + 20}{85 + 20 + 12 + 20 + 5} = 0.9501 \Rightarrow \boxed{Spec_{Setosa} = 95.01\%}$$

$$Spec_{Versi} = \frac{100 + 85 + 90 + 5}{100 + 85 + 90 + 5 + 18} = 0.9413 \Rightarrow \boxed{Spec_{Versi} = 94.13\%}$$

$$Spec_{Virgi} = \frac{100 + 90 + 115 + 18}{100 + 90 + 115 + 18 + 18} = 0.9371 \Rightarrow \boxed{Spec_{Virgi} = 93.71\%}$$

Evaluate classifier performance for individual classes.

e) Precision

$$Precision = \frac{TP}{TP + FP}$$

$$Prec_{Setosa} = \frac{100}{100 + 18} = \frac{100}{123} = 0.813 \Rightarrow \boxed{Prec_{Setosa} = 81.3\%}$$

$$Prec_{Versi} = \frac{85}{85 + 20 + 18} = \frac{85}{123} = 0.716 \Rightarrow \boxed{Prec_{Versi} = 71.6\%}$$

$$Prec_{Virgi} = \frac{90}{90 + 20} = \frac{90}{110} = 0.818 \Rightarrow \boxed{Prec_{Virgi} = 81.8\%}$$

Higher precision Setosa → fewer false positives