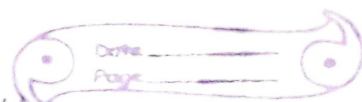


Rajat Sharma

11T2018141



Decision Tree problem

Problem 1 :- To create a decision tree with given impurity.

Solution :-

Taking first 8 entries as training data and next 8 entries as testing data from given dataset to explain the working of code

0 :- No , 1 :- Yes

AGE	HEART DISEASE	Sorting →	AGE	HEART DISEASE
63	0		37	No
67	1		41	No
67	1		56	No
37	0		57	No
41	0		62	Yes
56	0		63	No
62	1		67	Yes
57	0		67	yes

Next, I've made array of deciding factors which are just the collection of average age of two adjacent entries.

$$\text{factor 1} = \frac{37+41}{2} = \frac{78}{2} = 39$$

G.I. :- Gini Impurity

$$\boxed{\text{AGE} < 39}$$

HEART DISEASE

Yes	No
0	1

HEART DISEASE

Yes	No
3	4

$$\begin{aligned} \text{G.I.} &= 1 - \left(\frac{0}{1+0} \right)^2 - \left(\frac{1}{1+0} \right)^2 \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{G.I.} &= 1 - \left(\frac{3}{3+4} \right)^2 - \left(\frac{4}{3+4} \right)^2 \\ &= \frac{24}{49} \end{aligned}$$

G.I. for AGE < 39 is the weighted average of the impurities for the two leaves

$$\begin{aligned} &= \left(\frac{1}{1+7} \right) 0 + \left(\frac{7}{1+7} \right) \left(\frac{24}{49} \right) \\ &= 0.429 \end{aligned}$$

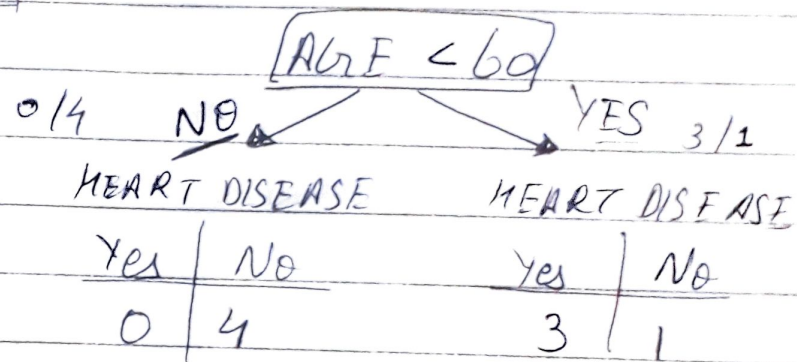
Similarly calculating for $\frac{41+56}{2}$, $\frac{56+57}{2}$,

$\frac{57+62}{2}$, $\frac{62+63}{2}$, $\frac{63+67}{2}$, $\frac{67+67}{2}$ we get

Factor	G.I.
39	0.429
49	0.375
57	0.3
60	0.187
63	0.370
65	0.208
67	0.208

→ smallest G.I.

so deciding factor would be 60



Testing data

AGE	HEART DISEASE	PREDICTION	
63	1	1	✓
53	1	0	✗
57	0	0	✓
56	0	0	✓
56	1	0	✗
44	0	0	✓
52	0	0	✓
57	0	0	✓

Accuracy = $\frac{6^3}{8^4}$

= 0.75

i.e. 75%