

Homework and lab - Part 1

Student	Study Hours	Pass (1) / Fail (0)
1	2	0
2	4	0
3	6	1
4	8	1
5	10	1

Step 1: Possible split points.

Between 2-4, 4-6, 6-8, 8-10, \Rightarrow 3, 5, 7, 9.

Step 2-3: calculate weighted Gini Impurity.

Split at 3:

- Left: $[0] \rightarrow \text{gini} = 0$
- Right: $[0, 1, 1, 1] \rightarrow \text{gini} = 1 - (1/4)^2 - (3/4)^2 = 0.375$
- Weighted: $(1/5) \times 0 + (4/5) \times 0.375 = 0.300$

Split at 5:

- Left: $[0, 0] \rightarrow \text{gini} = 0$
- Right: $[1, 1, 1] \rightarrow \text{gini} = 0$
- Weighted: $(2/5) \times 0 + (3/5) \times 0 = 0.000$

Split at 7:

- Left: $[0, 0, 1] \rightarrow \text{gini} = 1 - (2/3)^2 - (1/3)^2 = 0.444$
- Right: $[1, 1] \rightarrow \text{gini} = 0$
- Weighted: $(3/5) \times 0.444 + (2/5) \times 0 = 0.267$

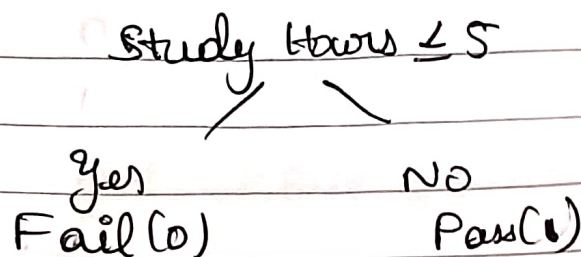
Split at 9:

- Left: $[0, 0, 1, 1] \rightarrow \text{gini} = 1 - (2/4)^2 - (2/4)^2 = 0.5$
- Right: $[1] \rightarrow \text{gini} = 0$
- Weighted: $(4/5) \times 0.5 + (1/5) \times 0 = 0.400$

Step 4: Best split

Split at 'S' has lowest Gini impurity (0.000)

Final Decision tree:



\therefore No further splitting needed since both child nodes are pure (Gini = 0)