

Required activity 9.3

Traffic: 14

No Traffic: 18

$$\text{Gini index of the parent node: } \frac{14}{32} * \frac{18}{32} + \frac{18}{32} * \frac{14}{32} = 0.4921875$$

Possible splits: [Weather, Day, Time]

1. Weather

Sunny: Traffic (4), No Traffic (17) = 21

Rainy: Traffic (10), No Traffic (1) = 11

Gini index of the weather split:

$$G(H1): \frac{4}{21} * \frac{17}{21} + \frac{17}{21} * \frac{4}{21} = 0.30839$$

$$G(H2): \frac{10}{11} * \frac{1}{11} + \frac{1}{11} * \frac{10}{11} = 0.1652$$

$$IG = G(\text{parent}) - [w1 * G(\text{child1}) + w2 * G(\text{child2})]$$

$$= 0.4921875 - (21/32 * 0.30839 + 11/32 * 0.1652)$$

$$= 0.23301$$

2. Day

Weekday: Traffic (7), No Traffic (13) = 20

Weekend: Traffic (7), No Traffic (5) = 12

Gini index of the day split:

$$G(H1): \frac{7}{20} * \frac{13}{20} + \frac{13}{20} * \frac{7}{20} = 0.455$$

$$G(H2): \frac{7}{12} * \frac{5}{12} + \frac{5}{12} * \frac{7}{12} = 0.4861$$

$$\begin{aligned} IG &= G(\text{parent}) - [w1 * G(\text{child1}) + w2 * G(\text{child2})] \\ &= 0.4921875 - (20/32 * 0.455 + 12/32 * 0.4861) \\ &= 0.0255 \end{aligned}$$

3. Time

8:00: Traffic (8), No Traffic (7) = 15

13:00: Traffic (6), No Traffic (11) = 17

Gini index of the time split:

$$G(H1): \frac{8}{15} * \frac{7}{15} + \frac{7}{15} * \frac{8}{15} = 0.4977$$

$$G(H2): \frac{6}{17} * \frac{11}{17} + \frac{11}{17} * \frac{6}{17} = 0.4567$$

$$\begin{aligned} IG &= G(\text{parent}) - [w1 * G(\text{child1}) + w2 * G(\text{child2})] \\ &= 0.4921875 - (15/32 * 0.4977 + 17/32 * 0.4567) \\ &= 0.01626 \text{ Not Viable} \end{aligned}$$

Based on the above information, we will split on **Weather** with maximum information gain.

After splitting on weather,

Possible splits: [Day, Time]

1. Day

a. Sunny:

Weekday: Traffic (2), No Traffic (13) = 15

Weekend: Traffic (2), No Traffic (4) = 6

Gini index of the Sunny-Day split:

$$G(H1): \frac{2}{15} * \frac{13}{15} + \frac{13}{15} * \frac{2}{15} = 0.231$$

$$G(H2): \frac{2}{6} * \frac{4}{6} + \frac{4}{6} * \frac{2}{6} = 0.4444$$

$$\begin{aligned} IG &= G(\text{sunny}) - [w1 * G(\text{child1}) + w2 * G(\text{child2})] \\ &= 0.30839 - (15/21 * 0.231 + 6/21 * 0.4444) \\ &= \mathbf{0.0164 \text{ Not Viable}} \end{aligned}$$

b. Rainy

Weekday: Traffic (5), No Traffic (0) = 5

Weekend: Traffic (5), No Traffic (1) = 6

Gini index of the Rainy-Day split:

$$G(H1): 0$$

$$G(H2): \frac{5}{6} * \frac{1}{6} + \frac{1}{6} * \frac{5}{6} = 0.2777$$

$$\begin{aligned} IG &= G(\text{rainy}) - [w1 * G(\text{child1}) + w2 * G(\text{child2})] \\ &= 0.1652 - (5/11 * 0.0 + 6/11 * 0.2777) \\ &= \mathbf{0.01372 \text{ Not viable}} \end{aligned}$$

2. Time

a. Sunny:

8:00: Traffic (3), No Traffic (6) = 9

13:00: Traffic (1), No Traffic (11) = 12

Gini index of the Sunny-Time split:

$$G(H1): \frac{3}{9} * \frac{6}{9} + \frac{6}{9} * \frac{3}{9} = 0.444$$

$$G(H2): \frac{1}{12} * \frac{11}{12} + \frac{11}{12} * \frac{1}{12} = 0.1527$$

$$\begin{aligned} IG &= G(\text{sunny}) - [w1 * G(\text{child1}) + w2 * G(\text{child2})] \\ &= 0.30839 - (9/21 * 0.444 + 12/21 * 0.1527) \\ &= 0.03084 \end{aligned}$$

b. Rainy

8:00: Traffic (5), No Traffic (1) = 6

13:00: Traffic (5), No Traffic (0) = 5

Gini index of the Rainy-Time split:

$$G(H1): \frac{5}{6} * \frac{1}{6} + \frac{1}{6} * \frac{5}{6} = 0.2777$$

$$G(H2): 0$$

$$\begin{aligned} IG &= G(\text{rainy}) - [w1 * G(\text{child1}) + w2 * G(\text{child2})] \\ &= 0.1652 - (6/11 * 0.2777 + 5/11 * 0.0) \\ &= 0.01372 \text{ Not Viable} \end{aligned}$$

From the above information, we can split the **sunny** branch by **Time** and **rainy** branch we can **terminate** since both of the IG are less than 0.02 and equal so there is no strong signal for either of those splits.

We can further split:

Possible splits: [Day]

1. Day

a. Sunny: 8:00:

Weekday: Traffic (2), No Traffic (6) = 8

Weekend: Traffic (1), No Traffic (0) = 1

Gini index of the Sunny-8:00 split:

$$G(H1): \frac{2}{8} * \frac{6}{8} + \frac{6}{8} * \frac{2}{8} = 0.375$$

$$G(H2): 0$$

$$\begin{aligned} IG &= G(\text{sunny-8:00}) - [w1 * G(\text{child1}) + w2 * G(\text{child2})] \\ &= 0.444 - (8/9 * 0.375 + 1/9 * 0) \\ &= 0.110 \end{aligned}$$

b. Sunny: 13:00:

Weekday: Traffic (0), No Traffic (7) = 7

Weekend: Traffic (1), No Traffic (4) = 5

Gini index of the Sunny-13:00 split:

$$G(H1): 0$$

$$G(H2): \frac{1}{5} * \frac{4}{5} + \frac{4}{5} * \frac{1}{5} = 0.32$$

$$IG = G(\text{sunny-13:00}) - [w1 * G(\text{child1}) + w2 * G(\text{child2})]$$

$$= 0.1527 - (7/12 * 0 + 5/12 * 0.32)$$

= 0.019 Not Viable

The resulting decision tree looks like.

