**Required activity 9.3**

Traffic: 14

No Traffic: 18

Gini index of the parent node: = 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): = 0.30839

G(H2): = 0.1652

IG = G(parent) – [ w1 \* G(child1) + w2 \* G(child2)]

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

= 0.23301

1. **Day**

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

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

Gini index of the day split:

G(H1): = 0.455

G(H2): = 0.4861

IG = G(parent) – [ w1 \* G(child1) + w2 \* G(child2)]

= 0. 4921875 - (20/32 \* 0.455 + 12/32 \* 0.4861)

= 0.0255

1. **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): = 0.4977

G(H2): = 0.4567

IG = G(parent) – [ w1 \* G(child1) + w2 \* G(child2)]

= 0. 4921875 - (15/32 \* 0.4977 + 17/32 \* 0.4567)

= 0.01626 Not Viable

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

After splitting on weather,

Possible splits: [Day, Time]

1. **Day**
2. Sunny:

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

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

Gini index of the Sunny-Day split:

G(H1): = 0.231

G(H2): = 0.4444

IG = G(sunny) – [ w1 \* G(child1) + w2 \* G(child2)]

= 0.30839- (15/21\*0.231+ 6/21 \* 0.4444)

= 0.0164 Not Viable

1. Rainy

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

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

Gini index of the Rainy-Day split:

G(H1):

G(H2): = 0.2777

IG = G(rainy) – [ w1 \* G(child1) + w2 \* G(child2)]

= 0.1652 - (5/11 \* 0.0 + 6/11 \* 0.2777)

= 0.01372 Not viable

1. **Time**
2. 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): = 0.444

G(H2): = 0.1527

IG = G(sunny) – [ w1 \* G(child1) + w2 \* G(child2)]

= 0.30839- (9/21\*0.444+ 12/21 \* 0.1527)

= 0.03084

1. 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): = 0.2777

G(H2):

IG = G(rainy) – [ w1 \* G(child1) + w2 \* G(child2)]

= 0.1652 - (6/11 \* 0.2777 + 5/11 \* 0.0)

= 0.01372 Not Viable

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**
2. 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): = 0.375

G(H2):

IG = G(sunny-8:00) – [ w1 \* G(child1) + w2 \* G(child2)]

= 0.444- (8/9\*0.375+ 1/9 \*0)

= 0.110

1. 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):

G(H2): = 0.32

IG = G(sunny-13:00) – [ w1 \* G(child1) + w2 \* G(child2)]

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

= 0.019 Not Viable

The resulting decision tree looks like.

Is it Sunny?

(Tr: 14, No Tr: 18)

GI: 0.4921

No Traffic

(Tr: 2, No Tr: 6)

GI: 0.375

Traffic

(Tr: 1, No Tr: 0)

GI: 0

No Traffic

(Tr: 1, No Tr: 11)

GI: 0.153

Traffic

(Tr: 10, No Tr: 1)

GI: 0.1652

Yes

No

Is it weekend?

(Tr: 3, No Tr: 6)

GI: 0.444

Yes

No

Is it morning?

(Tr: 4, No Tr: 17)

GI: 0.308

No

Yes