

### Decision Tree:-

1) Let's say we have a sample of 30 students with three variables

Gender (Boy/Girl)

Class (IX/X)

Height (5 to 6 ft)

15 out of 30 play Cricket in leisure time

		students	Play Cricket
Gender	Female	10	2
	Male	20	13
Height	< 5.5 ft	12	5
	≥ 5.5 ft	18	10
Class	IX	14	6
	X	16	9

Now, we want to create a model to predict who will play Cricket during leisure period.

1 Sunday ●

A) student = 30, play cricket = 15 (50%)

\* GINI:-

split on Gender:-

Female = 10 [play cricket = 2 (20%)]

male = 20 [play cricket = 13 (65%)]

$$\text{Gini Fox Female} = (2/10)^2 + (8/10)^2 = 0.68$$

$$\text{Gini Fox male} = (13/20)^2 + (7/20)^2 = 0.55$$

$$\text{Gini Fox Gender} = [(10/30) \times 0.68] + [(20/30) \times 0.55] \\ = 0.59 \checkmark$$

split on Height:-

< 5.5 ft = 12 [play cricket = 5 (42%)]

≥ 5.5 ft = 18 [play cricket = 10 (56%)]

$$\text{Gini Fox } < 5.5 \text{ ft} = (5/12)^2 + (7/12)^2 = 0.5128$$

$$\text{Gini Fox } \geq 5.5 \text{ ft} = (10/18)^2 + (8/18)^2 = 0.5072$$

$$\text{Gini Fox height} = [(12/30) \times 0.5128] + [(18/30) \times 0.5072] \\ = 0.5094$$

split on class:-

class 1x = 14 [play cricket = 6 (43%)]

class x = 16 [play cricket = 9 (56%)]

$$\text{Gini Fox class 1x} = (6/14)^2 + (8/14)^2 = 0.5098$$

$$\text{Gini Fox class x} = (9/16)^2 + (7/16)^2 = 0.5072$$

$$\text{Gini Fox class} = [(14/30) \times 0.5098] + [(16/30) \times 0.5072] \\ = 0.5084$$

Gini Fox Gender is higher, hence the node split will take place on Gender.

Choose a job you love, and you will never have to work a day in your life. - Confucius





\* INFORMATION GAIN:- Entropy for playing = 1

Split on Gender:-

$$\text{Entropy for female} = -\frac{2}{10} \log_2\left(\frac{2}{10}\right) - \frac{8}{10} \log_2\left(\frac{8}{10}\right) = 0.7219$$

$$\text{Entropy for male} = -\frac{13}{20} \log_2\left(\frac{13}{20}\right) - \frac{7}{20} \log_2\left(\frac{7}{20}\right) = 0.9342$$

Information Gain for gender

$$= 1 - \left[ \frac{10}{30} \times 0.7219 + \frac{20}{30} \times 0.9342 \right]$$

$$= 0.1366$$

Split on Height:-

$$\text{Entropy for } < 5.5 \text{ ft} = -\frac{5}{12} \log_2\left(\frac{5}{12}\right) - \frac{7}{12} \log_2\left(\frac{7}{12}\right) = 0.9799$$

$$\text{Entropy for } \geq 5.5 \text{ ft} = -\frac{10}{18} \log_2\left(\frac{10}{18}\right) - \frac{8}{18} \log_2\left(\frac{8}{18}\right) = 0.9912$$

Information Gain for gender =

$$= 1 - \left[ \frac{12}{20} \times 0.9799 + \frac{18}{30} \times 0.9912 \right]$$

$$= 0.0135$$

Split on class:-

$$\text{Entropy for class 1X} = -\frac{6}{14} \log_2\left(\frac{6}{14}\right) - \frac{8}{14} \log_2\left(\frac{8}{14}\right) = 0.9853$$

$$\text{Entropy for class X} = -\frac{9}{16} \log_2\left(\frac{9}{16}\right) - \frac{7}{16} \log_2\left(\frac{7}{16}\right) = 0.9887$$

8 9 10 11 12 13 14  
15 16 17 18 19 20 21  
22 23 24 25 26 27 28  
29 30 31



December 2019



Thursday 12

346-19 / Week 50

\* INFORMATION GAIN:- Entropy Fox playing = 1

Split on Gender:-

$$\text{Entropy Fox female} = -\frac{2}{10} \log_2\left(\frac{2}{10}\right) - \frac{8}{10} \log_2\left(\frac{8}{10}\right) = 0.7219$$

$$\text{Entropy Fox male} = -\frac{13}{20} \log_2\left(\frac{13}{20}\right) - \frac{7}{20} \log_2\left(\frac{7}{20}\right) = 0.9342$$

Information Gain Fox gender

$$= 1 - \left[ \frac{10}{30} \times 0.7219 + \frac{20}{30} \times 0.9342 \right]$$

$$= 0.1366$$

Split on Height:-

$$\text{Entropy Fox } < 5.5 \text{ ft} = -\frac{5}{12} \log_2\left(\frac{5}{12}\right) - \frac{7}{12} \log_2\left(\frac{7}{12}\right) = 0.9799$$

$$\text{Entropy Fox } \geq 5.5 \text{ ft} = -\frac{10}{18} \log_2\left(\frac{10}{18}\right) - \frac{8}{18} \log_2\left(\frac{8}{18}\right) = 0.9912$$

Information Gain Fox gender =

$$= 1 - \left[ \frac{12}{20} \times 0.9799 + \frac{18}{30} \times 0.9912 \right]$$

$$= 0.0135$$

Split on class:-

$$\text{Entropy Fox class 1X} = -\frac{6}{14} \log_2\left(\frac{6}{14}\right) - \frac{8}{14} \log_2\left(\frac{8}{14}\right) = 0.9853$$

$$\text{Entropy Fox class X} = -\frac{9}{16} \log_2\left(\frac{9}{16}\right) - \frac{7}{16} \log_2\left(\frac{7}{16}\right) = 0.9887$$

8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

Good friends, good books and a sleepy conscience; this is the ideal life. - Mark Twain





Information gain for class =

$$= 1 - \left[ \frac{14}{30} \times 0.9853 + \frac{16}{30} \times 0.9887 \right]$$

$$= 0.0129$$

Information Gain for genders is highest, hence the node split will take place on genders

Information Gain =

Entropy of parent - weighted entropy of subnodes

Gini  $\rightarrow$  For Numeric

Information Gain  $\rightarrow$  Categorical

2) Purchase Data set:- Information Gain

Entropy for purchase:-

$$= -\frac{24}{30} \log_2\left(\frac{24}{30}\right) - \frac{6}{30} \log_2\left(\frac{6}{30}\right)$$

$$= 0.7219$$

split on Holiday =

No:- 11 [Purchase yes=9, No=2]

Yes:- 19 [Purchase yes=15, No=4]

$$\text{Entropy for No holiday} = -\frac{9}{11} \log_2\left(\frac{9}{11}\right) - \frac{2}{11} \log_2\left(\frac{2}{11}\right)$$

$$= 0.684$$

$$\text{Entropy for yes holiday} = -\frac{15}{19} \log_2\left(\frac{15}{19}\right) - \frac{4}{19} \log_2\left(\frac{4}{19}\right)$$

$$= 0.742$$

In three words I can sum up everything I've learned about life: it goes on. - Robert Frost



December 2019



Saturday 14

148-17 / Week 50

Information gain for holiday

$$= 0.7219 - \left[ \frac{11}{30} \times 0.684 + \frac{19}{30} \times 0.742 \right]$$
$$= 0.0009$$

Split on Discount:-

No:-10 [Purchase yes=5, No=5]

Yes:-20 [Purchase yes=19, No=1]

Entropy for No discount =  $-\frac{5}{10} \log_2\left(\frac{5}{10}\right) - \frac{5}{10} \log_2\left(\frac{5}{10}\right)$

$$= 1$$

Entropy for yes discount =  $-\frac{19}{20} \log_2\left(\frac{19}{20}\right) - \frac{1}{20} \log_2\left(\frac{1}{20}\right)$

$$= 0.286$$

Information gain for discount

$$= 0.721 - \left[ \frac{10}{30} \times 1 + \frac{20}{30} \times 0.286 \right]$$

$$= 0.197 \checkmark$$

Split on Free delivery:-

No=7 [Purchase yes=3, No=4]

Yes=23 [Purchase yes=21, No=2] Sunday

Entropy for No free delivery =  $-\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right)$

$$= 0.985$$

Entropy for free delivery =

$$= -\frac{21}{23} \log_2\left(\frac{21}{23}\right) - \frac{2}{23} \log_2\left(\frac{2}{23}\right)$$

$$= 0.426$$

December 2019						
S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

Tomorrow hopes we have learned something from yesterday. - John Wayne



16 Monday

ASD-18 / Week 50

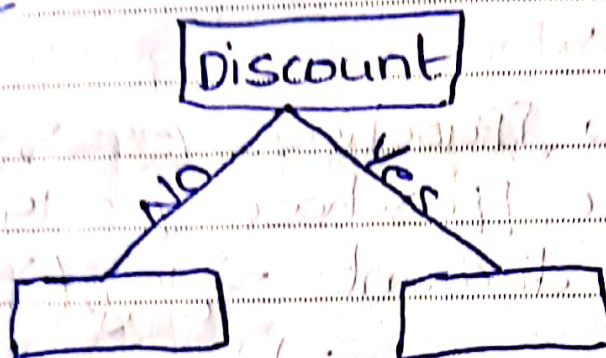


December 2019

Information gain for free delivery =

$$= 0.721 - \left[ \frac{5}{30} \times 0.985 + \frac{23}{30} \times 0.426 \right]$$

$$= 0.164$$



Discount No :-

Entropy for No Discount

$$= -\frac{5}{10} \log_2 \left( \frac{5}{10} \right) - \frac{5}{10} \log_2 \left( \frac{5}{10} \right)$$
$$= 1$$

split on holiday :-

No = 2 [Purchase yes = 0, No = 2]

Yes = 8 [Purchase yes = 5, No = 3]

$$\text{Entropy for No holiday} = -\frac{0}{2} \log_2 \left( \frac{0}{2} \right) - \frac{2}{2} \log_2 \left( \frac{2}{2} \right)$$
$$= 0$$

$$\text{Entropy for yes holiday} = -\frac{5}{8} \log_2 \left( \frac{5}{8} \right) - \frac{3}{8} \log_2 \left( \frac{3}{8} \right)$$
$$= 0.954432$$

Information gain for holiday

$$= 1 - \left[ 0 + \frac{8}{10} \times 0.954432 \right]$$

$$= 0.23645$$



December 2019



Tuesday 17

351-14 / Week 51

split on Free Delivery:-

No=5 [Purchase Yes=1, No=4]

Yes=5 [Purchase Yes=4, No=1]

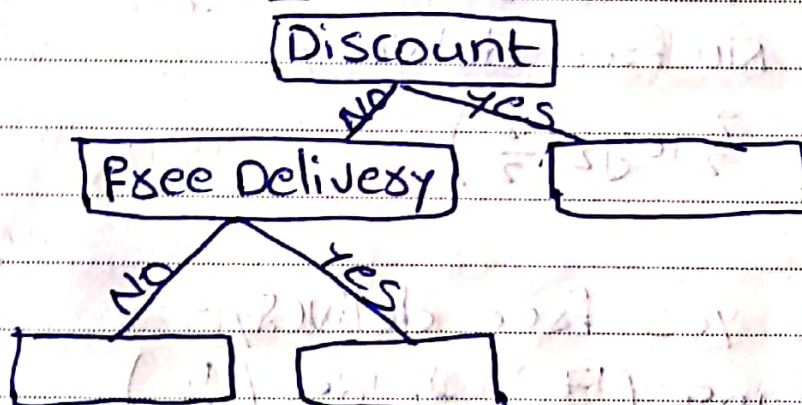
$$\text{Entropy for } x \text{ free delivery} = -\frac{1}{5} \log_2\left(\frac{1}{5}\right) - \frac{4}{5} \log_2\left(\frac{4}{5}\right) = 0.721928$$

$$\text{Entropy for yes free delivery} = -\frac{4}{5} \log_2\left(\frac{4}{5}\right) - \frac{1}{5} \log_2\left(\frac{1}{5}\right) = 0.721928$$

Information gain for free delivery

$$= 1 - \left[ \frac{5}{10} \times 0.721928 + \frac{5}{10} \times 0.721928 \right]$$

$$= 0.278072 \checkmark$$



Discount yes:-

Entropy for yes discount =

$$= -\frac{19}{20} \log_2\left(\frac{19}{20}\right) - \frac{1}{20} \log_2\left(\frac{1}{20}\right)$$

$$= 0.286396$$

split on Holiday:-

No:-9 [Purchase Yes=9, No=0]

Yes:-11 [Purchase Yes=10, No=1]

December 2019						
S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				





$$\text{Entropy for No holiday} = -\frac{9}{9} \log_2\left(\frac{9}{9}\right) - 0$$

$$= 0$$

$$\text{Entropy for yes holiday} = -\frac{10}{11} \log_2\left(\frac{10}{11}\right) - \frac{1}{11} \log_2\left(\frac{1}{11}\right)$$

$$= 0.439496$$

Information gain for holiday

$$= 0.286 - \left[ \frac{9}{20} \times 0 + \frac{11}{20} \times 0.439 \right]$$

$$= 0.04 \checkmark$$

Split on free delivery:-

No=2 [Purchase yes=2, No=0]

Yes=18 [Purchase yes=17, No=1]

Entropy for No free delivery =

$$= 0 - \frac{2}{2} \log_2\left(\frac{2}{2}\right)$$

$$= 0$$

Entropy for yes free delivery =

$$= -\frac{17}{18} \log_2\left(\frac{17}{18}\right) - \frac{1}{18} \log_2\left(\frac{1}{18}\right)$$

$$= 0.309543$$

Information gain for free delivery

$$= 0.286396 - \left[ 0 + \frac{18}{20} \times 0.309543 \right]$$

$$= 0.007$$



