YEAR 2022

JANUARY
Su	Mo	Tu	We	Th	Fr	Sa
30	31					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

FEBRUARY
Su	Mo	Tu	We	Th	Fr	Sa
		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					

MARCH
Su	Mo	Tu	We	Th	Fr	Sa
		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		

APRIL
Su	Mo	Tu	We	Th	Fr	Sa
					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

MAY
Su	Mo	Tu	We	Th	Fr	Sa
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				

JUNE
Su	Mo	Tu	We	Th	Fr	Sa
			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		

JULY
Su	Mo	Tu	We	Th	Fr	Sa
31					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

AUGUST
Su	Mo	Tu	We	Th	Fr	Sa
	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			

SEPTEMBER
Su	Mo	Tu	We	Th	Fr	Sa
				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	

OCTOBER
Su	Mo	Tu	We	Th	Fr	Sa
30	31					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

NOVEMBER
Su	Mo	Tu	We	Th	Fr	Sa
		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			

DECEMBER
Su	Mo	Tu	We	Th	Fr	Sa
				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

YEAR 2023

JANUARY
Su	Mo	Tu	We	Th	Fr	Sa
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				

FEBRUARY
Su	Mo	Tu	We	Th	Fr	Sa
			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				

MARCH
Su	Mo	Tu	We	Th	Fr	Sa
			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	

APRIL
Su	Mo	Tu	We	Th	Fr	Sa
30						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

MAY
Su	Mo	Tu	We	Th	Fr	Sa
	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			

JUNE
Su	Mo	Tu	We	Th	Fr	Sa
				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	

JULY
Su	Mo	Tu	We	Th	Fr	Sa
30	31					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

AUGUST
Su	Mo	Tu	We	Th	Fr	Sa
		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		

SEPTEMBER
Su	Mo	Tu	We	Th	Fr	Sa
					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

OCTOBER
Su	Mo	Tu	We	Th	Fr	Sa
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				

NOVEMBER
Su	Mo	Tu	We	Th	Fr	Sa
			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		

DECEMBER
Su	Mo	Tu	We	Th	Fr	Sa
31					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

YEAR 2024

JANUARY
Su	Mo	Tu	We	Th	Fr	Sa
	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			

FEBRUARY
Su	Mo	Tu	We	Th	Fr	Sa
				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		

MARCH
Su	Mo	Tu	We	Th	Fr	Sa
31					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

APRIL
Su	Mo	Tu	We	Th	Fr	Sa
	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				

MAY
Su	Mo	Tu	We	Th	Fr	Sa
			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	

JUNE
Su	Mo	Tu	We	Th	Fr	Sa
30						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

JULY
Su	Mo	Tu	We	Th	Fr	Sa
	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			

AUGUST
Su	Mo	Tu	We	Th	Fr	Sa
				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

SEPTEMBER
Su	Mo	Tu	We	Th	Fr	Sa
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					

OCTOBER
Su	Mo	Tu	We	Th	Fr	Sa
		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		

NOVEMBER
Su	Mo	Tu	We	Th	Fr	Sa
					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

DECEMBER
Su	Mo	Tu	We	Th	Fr	Sa
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				

1.

(a)

Color

#28, #3, #34, #32 (YES)
#1, #26, #11, #13 (NO)

White

black

#28, (YES)
#11, #13 (NO)

#3, #34, #32 (YES)
#1, #26, (NO)

$$\text{Info}(D) = -\frac{4}{8}\log_2\frac{4}{8} - \frac{4}{8}\log_2\frac{4}{8} = 1$$

$$\text{Info}(\text{Color}) = \frac{3}{8}\left(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2 2/3\right) + \frac{5}{8}\left(-\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2 2/5\right)$$

$$= 0.3431 + 0.6068 = 0.95$$

$$\text{Information gain}(\text{Color}) = 1 - 0.95 = 0.05$$

Size

#28, #3, #34, #32
#1, #26, #11, #1

large

Small

#28,
#1, #26, #13

#3, #34, #32
#11,

$$\text{Info}(\text{Size}) = \frac{4}{8}\left(-\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4}\right)$$

$$+ \frac{4}{8}\left(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}\right)$$

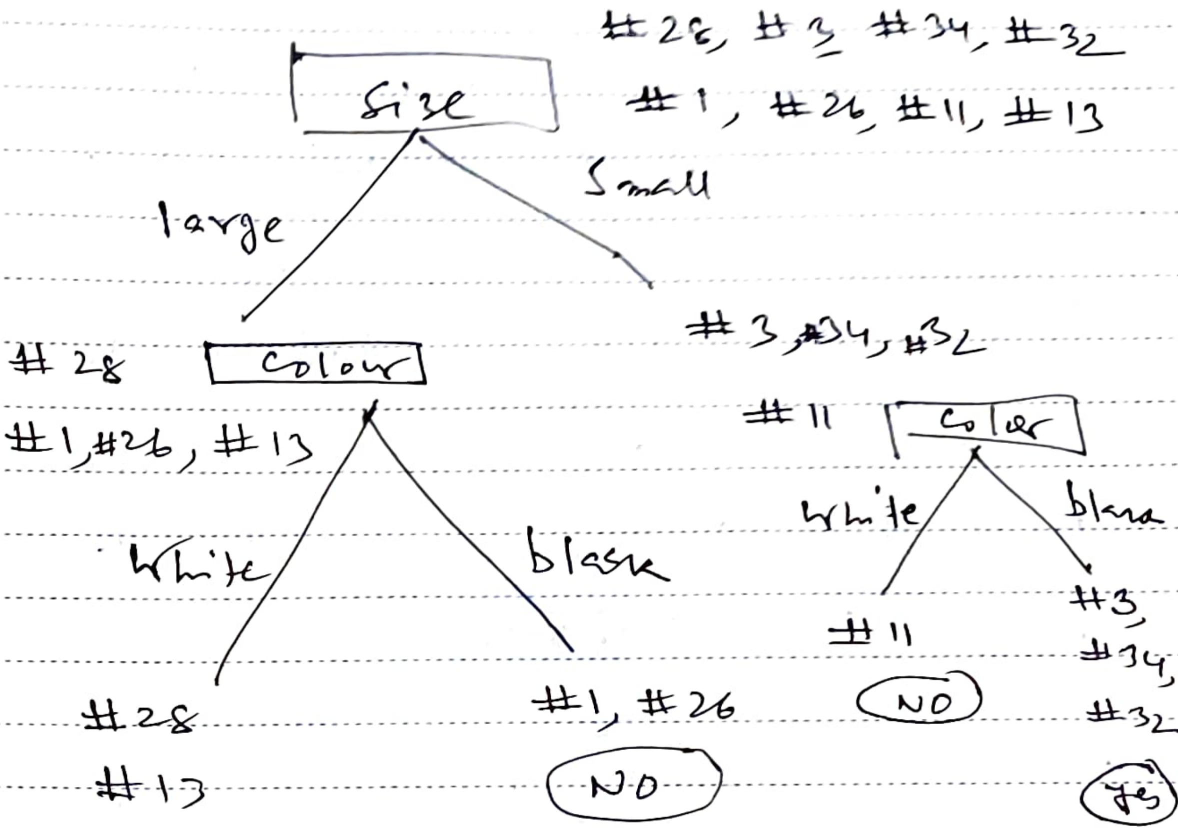
$$= 0.56$$

$$\text{Information gain}(\text{Size}) = 1 - 0.56$$
$$= 0.44$$

Since information gain of Size is larger than that of colour, Size should be used as the root of the tree. Now we have to construct a decision tree using Size and colour attribute. We should use Size as the root.

Sat Sun Mon Tue Wed Thu Fri

Date :



b.

Unique-id will have 8 branchs and each branch will contain a single yes or No example. This feature will have the highest information gain because the examples have been separated. The unique-id need to be selected as the root of the tree. But this tree will have no predictive power because

Sat Sun Mon Tue Wed Thu Fri

Date :

it does has no correlation with unique-id.
thus this type of attribute should be avoided in constructing the decision tree. This type of attribute is automatically avoided if Gain Ratio is used instead of information gain to select the attributes.

(c) tail-length is a numeric attribute. We can discretize it before we start constructing the tree. We can convert tail-length values into two categories, one is large and the other is small. If we choose, for example, 2.5 is the cut-off value, then 5.6, 3.8, 4.2, 3.5 will be converted to large and the other will be converted to small.

d

X = Six events each with probability 1/6

Y = two events each with probability 1/2

$$\text{Entropy}(X) = \cancel{-\frac{1}{6}\log_2\frac{1}{6}} \cancel{-\frac{1}{6}\log_2\frac{1}{6}} \cancel{-\frac{1}{6}\log_2\frac{1}{6}} \cancel{-\frac{1}{6}\log_2\frac{1}{6}} \cancel{-\frac{1}{6}\log_2\frac{1}{6}} \cancel{-\frac{1}{6}\log_2\frac{1}{6}}$$

$$\cancel{= -\log_2\frac{1}{6}} = -\frac{1}{6}\log_2\frac{1}{6} - \frac{5}{6}\log_2\frac{5}{6}$$

$$= 0.65$$

$$\text{Entropy}(Y) = -\frac{1}{2}\log_2 1/2 - \frac{1}{2}\log_2 1/2$$

$$= 1$$

$$\text{Entropy}(Y) > \text{Entropy}(X)$$

2.

$x_1 = 1 \quad x_2 = a \quad x_3 = v$

$P(Y=0 \mid x_1=1, x_2=a, x_3=v)$

$= P(x_1=1, x_2=a, x_3=v \mid Y=0) \times P(Y=0)$

$= P(x_1=1 \mid Y=0) \times P(x_2=a \mid Y=0)$

$\times P(x_3=v \mid Y=0) \times P(Y=0)$

$= \frac{1}{\cancel{4}_2} \times \frac{\cancel{2}}{4} \times \frac{1}{\cancel{4}} \times \frac{\cancel{4}}{9} = \frac{1}{72}$

$P(Y=1 \mid x_1=1, x_2=a, x_3=v)$

$= P(x_1=1 / Y=1) \; P(x_2=a \mid Y=1) \; P(x_3=v \mid Y=1)$

$P(Y=1)$

$= \frac{1}{5} \times \frac{1}{5} \times \frac{1}{\cancel{5}} \times \frac{\cancel{5}}{9} = \frac{1}{225}$

[illegible]

$\therefore$ Most likely value of $Y = 0$

3.

(a) $X = \{A, B, C\}$

$P(A) = 0.5 \quad P(B) = 0.3$

$\therefore P(C) = 1 - 0.5 - 0.3 = 0.2$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= 0.5 + 0.3 - 0.5 \times 0.3 = 0.8 - 0.15$

$= 0.65$

(b) $P(A) =$ H,H OR H,T out of HH, HT, TH and TT

$= \frac{2}{4} = \frac{1}{2}$

$P(B) =$ TH or TT $= \frac{2}{4} = \frac{1}{2}$

$P(A \cap B) = \frac{1}{2} \times \frac{1}{2}$
$= 1/4$

(c) (i) $P(B|A) = 0.5$

$P(B) = 0.1 + x + 0.2 + 0.1 = 0.4 + x$

For absolute independence

$P(B|A) = P(B)$

$\Rightarrow 0.4 + x = 0.5$

$\Rightarrow x = 0.1$

(ii) $P(B) = 0.5 = P(B|A)$

$\therefore P(B') = 1 - 0.5 = 0.5$

$P(B') = 0.2 + 0.1 + y + 0.1 = 0.5$

$\therefore y = 0.1$

(iii) $P(A \mid B \wedge C)$

$$= \frac{P(A \wedge B \wedge C)}{P(B \wedge C)} = \frac{0.1}{0.1 + 0.2} = \frac{0.1}{0.3}$$

$$= 0.33$$

4.

(i) Entropy (D)

$$= \text{Entropy}(9P, 7N)$$

$$= -\frac{9}{16}\log_2 \frac{9}{16} - \frac{7}{16}\log_2 \frac{7}{16}$$

(ii) As usual. calculate

(iii) Based on information gain, if information gain of color is greater than that of ~~color~~ size, then

$$P(\text{Pandemic} \wedge \text{Public univ} \mid \text{online class})$$

$$= P(\text{Pandemic} \mid \text{online class}) \times P(\text{Public univ} \mid \text{online class})$$

if conditional independency holds.

$$P(\text{Pandemic} \wedge \text{Public univ} \mid \text{online class})$$

$$= \frac{P(\text{Pandemic} \wedge \text{Public univ} \wedge \text{online class})}{P(\text{online class})}$$

$$= \frac{0.142}{0.142 + 0.103 + 0.165 + 0.217} = \frac{0.142}{0.627} = 0.23$$

$$P(\text{Pandemic} \mid \text{online class})$$

$$= \frac{P(\text{Pandemic} \wedge \text{Online class})}{P(\text{online class})}$$

$$= \frac{0.142 + 0.103}{0.142 + 0.103 + 0.165 + 0.217} = \frac{0.245}{0.627} = \cancel{0.39}$$

$$= 0.39$$

$$P(\text{Public univ} \mid \text{online class})$$

$$= \frac{P(\text{Public univ} \wedge \text{online class})}{P(\text{online class})}$$

$$= \frac{0.142 + 0.165}{0.627} = 0.49$$

Since $0.23 \neq 0.39 \times 0.49$, the statement is false.

S. M. PLASTIC INDUSTRIES LTD.

Sat Sun Mon Tue Wed Thu Fri

Date

(ii)

P (prior win | online class)

= P (prior win) if they are independent.

~~P (prior win | online class)~~

$$= \frac{P(\text{prior win}) \wedge P(\text{online class})}{P(\text{online class})}$$

$$= \frac{(0.103 + 0.146 + 0.217 + 0.118)(0.142 + 0.103 + 0.165 + 0.217)}{0.627}$$

(calculated earlier)

$\cancel{0.584 \times 0.627}$

$=$

$$P(\text{priv univ} \mid \text{online class})$$

$$= \frac{P(\text{priv univ} \wedge \text{online class})}{P(\text{online class})}$$

$$= \frac{0.103 + 0.217}{0.627} = \cancel{0.45}\ 0.51$$

$$P(\text{priv univ}) = 0.103 + 0.146 + 0.217 + 0.118$$

$$= 0.584$$

Statement is false

(iii) $P(\text{offline clas})$

$= 0.037 + 0.146 + 0.072 + 0.118$

$= 0.373$

(iv) $P(\text{Priv univ} \wedge \text{Offline clas} \mid \text{No Pandemic})$

$$= \frac{P(\text{Priv univ} \wedge \text{Offline clas} \wedge \text{No Pandemic})}{P(\text{No Pandemic})}$$

$$= \frac{0.118}{0.165 + 0.072 + 0.217 + 0.118} = 0.21$$

v) Naive Bayes is called Naive Bayes because adopting the strong assumption that attributes are independent. It outperforms bayes theorem because it avoids calculating the denominator of the Bay's rule.

(vi) $$P(yes|X) = \frac{P(x|yes) \cdot P(yes)}{P(x)}$$

$$P(No|x) = \frac{P(x|No) \cdot P(No)}{P(x)}$$

Since we need to compare $P(yes|x)$ and $P(No|x)$, we can omit $P(x)$ from the right hand side of both the equations.

weight value can be replaced by

12 (heavy) 8 (light) 17 (heavy)

13 (heavy) 4 (light) 5 (light)

18 (heavy) 11 (heavy) 9 (light)

6 (light)

Thus numeric attribute weight is converted to discrete attribute having only two values heavy and light.

(a) ~~P(dog ∧ heavy)~~

P(dog ∧ heavy | black)

If dog is conditionally independent of heavy given black then the following condition holds:

$P(dog \wedge heavy \mid color = black)$

$= P(dog \mid black)\ P(heavy \mid black)$

Now $P(dog \wedge heavy \mid color = black)$

$= 2/3$

[3 black animals out of which 2 are dogs and heavy]

$P(\text{dog} \mid \text{black}) = \frac{2}{3}$ (2 are dogs out of 3 black animals)

$P(\text{heavy} \mid \text{black})$

$= 2/3$ (2 are heavy out of 3 black animals)

Here $\frac{2}{3} \neq \frac{2}{3} \times \frac{2}{3}$

So the statement is false.

(D) $P(\underset{=\text{yes}}{\text{pet}} \mid \text{orange} \wedge \text{cat})$

$= P(\text{orange} \wedge \text{cat} \mid \underset{=\text{yes}}{\text{pet}}) \times P(\underset{=\text{yes}}{\text{pet}})$

$= P(\text{orange} \mid \text{Pet} = \text{Yes}) \times P(\text{cat} \mid \text{Pet} = \text{Yes}) \times P(\text{Pet} = \text{Yes})$

$= \frac{\cancel{2}}{\cancel{6}\,3} \times \frac{\cancel{2}}{\cancel{4}} \times \frac{\cancel{6}}{\cancel{10}\,5} = \frac{1}{15}$

$P(\text{Pet} = \text{No} \mid \text{orage a cat})$

$= P(\text{orange} \mid \text{Pet} = \text{No}) \times P(\text{cat} \mid \text{Pet} = \text{No}) \times P(\text{Pet} = \text{No})$

$= \frac{1}{4} \times \frac{1}{\cancel{4}} \times \frac{\cancel{4}}{10} = \frac{1}{40}$

$\therefore$ It is more likely that an orange cat is a pet animal.

7.

(A) Find information gain of Alpha and Beta and decide the better Attribute is the one which has more information gain. (Do it yourself)

(D)

A H T T T T

B H T T - -

The 4th and fifth outcome could be HH HT TH or TT
For every one we calculate the entropy and take the combination which has maximum entropy.

B can be as follows

HTTHH $\quad -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}$

$= 0.971$

HTTHT $\quad -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}$

$= 0.971$

HTTTH $\quad -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2 3/5$

$= 0.971$

HTTTT $\quad -\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2 4/5$

$= 0.721$

For A

A HTTTT $= -\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2 4/5$

$= 0.721$

So HTTHH, HTTHT and HTTTH (B) have high impurity.

(c) If gamma is continuous variable, we have to discretize it before constructing the decision tree.

8. $P(\text{Positive} \mid \text{test} = \text{Positive}) = 0.85$

$P(\text{Positive} \mid \text{test} = \text{Negative}) = 0.10$

(a) $P(\text{Positive}) = 0.7 \quad \therefore P(\text{Negative}) = 0.3$

$P(\text{test} = \text{Positive})$

$= P(\text{test} = \text{Positive} \wedge \text{Positive})$

$+ P(\text{test} = \text{Positive} \wedge \text{Negative})$

$$= P(\text{Positive} \mid \text{test} = \text{Positive}) \cdot P(\text{test} = \text{Positive})$$

$$+ P(\text{Negative} \mid \text{test} = \text{Positive}) \cdot P(\text{test} = \text{Positive})$$

$$= 0.85 \times$$

$$= P(\text{test} = \text{Positive} \mid \text{Positive}) \cdot P(\text{Positive})$$

$$+ P(\text{test} = \text{Positive} \mid \text{Negative}) \cdot P(\text{Negative})$$

$$= \frac{P(\text{Positive} \mid \text{test} = \text{Positive}) \cdot P(\text{Positive})}{P(\text{Positive})} \cdot P(\text{Positive})$$

$$+ P(\text{Negative} \mid \text{test} = \text{Positive})\, P($$

8.

(b)

$P(\text{covid} \mid \text{test} = \text{Positive}) = 0.85$

$P(\neg\text{covid} \mid \text{test} = \text{Positive}) = 0.15$

$P(\text{covid} \mid \text{test} = \text{Negative}) = 0.10$

$P(\neg\text{covid} \mid \text{test} = \text{Negative}) = 0.90$

$P(\cancel{\text{covid}}) = \cancel{0.70}$ $P(\text{test} = \text{Positive}) = 0.70$

$P(\text{test} = \text{Positive} \mid \text{covid})$

$$= \frac{P(\text{covid} \mid \text{test} = \text{Positive}) \times P(\text{test} = \text{Positive})}{P(\text{covid})}$$

$$P(a|b) = \frac{P(b|a) \cdot P(a)}{P(b)}$$

$$= \frac{0.85 \times 0.70}{P(covid)} = \frac{0.6}{\alpha} = \frac{0.6}{0.6 + 0.03}$$

$$= 0.95$$

$$P(test = Negative \mid covid)$$

$$= \frac{P(covid \mid test = Negative) \cdot P(test = Negative)}{P(covid)}$$

$$= \frac{0.10 \times 0.30}{P(covid)} = \frac{0.03}{\alpha} = \frac{0.03}{0.6 + 0.03}$$

$$= 0.05$$

test = Positive more likely

(2)

$P(\text{Covid} \mid \text{Test} = \text{positive}) = 0.85$

$P(\neg\text{Covid} \mid \text{Test} = +ve) = 0.15$

$P(\text{Covid} \mid \text{Test} = -ve) = 0.10$

$P(\neg\text{Covid} \mid \text{Test} = -ve) = 0.90$

$P(\text{Covid}) = 0.70$

$P(\text{Test} = +ve) = ?$

$$P(\text{Test} = +ve) = P(\text{Test} = +ve \wedge \text{Covid}) + P(\text{Test} = +ve \wedge \neg\text{Covid})$$

$$= P(\text{covid} \mid \text{Test} = +ve)$$

$$= P[\text{test} = +ve \mid \text{covid}) \cdot P(\text{covid})$$

$$+ P(\text{test} = +ve \mid \neg\text{covid})\ P(\neg\text{covid})$$

$$= 0.70\ P(\text{test} = +ve \mid \text{covid})$$

$$+ 0.30\ (\text{test} = +ve \mid \neg\text{covid})$$

9.

(a) $$P(\text{Rugby} \mid \text{male}) = \frac{P(\text{Rugby} \wedge \text{male})}{P(\text{Male})}$$

$$= \frac{0.2}{0.54}$$

(b) $$P(\text{female} \mid \text{Football}) = \frac{P(\text{female} \wedge \text{football})}{P(\text{football})}$$

$$= \frac{0.15}{0.39}$$

(c) $$P(\text{Football} \vee \text{other})$$

$$= P(\text{football}) + P(\text{other}) - P(\text{football} \wedge \text{other})$$

$$= 0.39 + 0.36 - \cancel{0.39} \times \cancel{0.56}\ (\cancel{1 - 0.45})$$

$$0.39 \times 0.36$$

(d)

$$P(\text{Rugby} \mid \text{Female})$$

$$= \frac{P(\text{Rugby} \wedge \text{Female})}{P(\text{Female})}$$

$$= \frac{0.05}{0.46} = 0.11$$

$$P(\text{Rugby}) = 0.25$$

Since $P(\text{Rugby} \mid \text{Female}) \neq P(\text{Rugby})$

therefore playing Rugby depends on females.

10.

(a) Find information gain of x1, x2 and x3 and select that one which has the maximum information gain

(b) No, ID3 will not be suitable for this classification task. We have to convert continuous attributes into discrete attributes before we start building the tree.

(c) 10 Male → Smoker / Non-smoker = 70 : 30

5 Female → Smoker / Non-smoker = 20 : 80

~~Entropy~~ P (smoking | male)

$$= \frac{10}{15} \times \frac{70}{100} = \frac{7}{15} = \cancel{0.30} \; 0.47$$

Entropy (smoking | male)

$= - p \log_2 p$

$= - 0.47 \log_2 0.47 = 0.52$

11.

(a) $P(C=1, M=0, Y=1)$

$= P(C=1, M=0 \mid Y=1) \cdot P(Y=1)$

$= P(C=1 \mid y=1) \; P(M=0 \mid Y=1) \cdot P(Y=1)$

$= P_2 \times (1-P_4) \cdot (1-q)$

(b) $P(C=1, M=0, Y=0)$

$= P(C=1, M=0 \mid Y=0) \; P(Y=0)$

$$= P(C=1 \mid Y=0)\; P(M=0 \mid Y=0)\; P(Y=0)$$

$$= (1-P_1)\,(1-P_3)\,q$$

(c) $P(Y=1 \mid C=1, M=0)$

$$= \frac{P(Y=1, C=1, M=0)}{P(C=1, M=0)}$$

$$= \frac{P_2 \times (1-P_4) \times (1-q)}{P(C=1, M=0, Y=1) + P(C=1, M=0, Y=0)} \quad [\text{from } ⓐ]$$

$$= \frac{P_2 \times (1-P_4) \times (1-q)}{P_2 \times (1-P_4) \times (1-q) + (1-P_1) \times (1-P_3) \times q}$$

[from ⓐ and ⓑ and marginilization]

Sat Sun Mon Tue Wed Thu Fri

Date ....

(f) From the table

$q = P(Y=0) = \frac{3}{10} = 0.3$

$P_1 = P(C=0 \mid Y=0) = \frac{2}{3}$

$P_2 = P(C=1 \mid Y=1) = \frac{4}{7}$

$P_3 = P(M=1 \mid Y=0) = \frac{1}{3}$

$P_4 = P(M=1 \mid Y=1) = \frac{2}{7}$

(e) $P(Y=1 \mid C=1, M=0)$

$= P(M=0 \mid Y=1) \cdot P(C=1 \mid Y=1) \cdot P(Y=1)$

$= \left(1-\frac{2}{7}\right)\left(1-\frac{4}{7}\right)(1-0.3) = 0.214$

$$P(Y=0 \mid C=1, M=0)$$

$$= P(C=1 \mid Y=0) \cdot P(M=0 \mid Y=0) \cdot P(Y=0)$$

$$= \left(1-\frac{2}{3}\right)\left(1-\frac{1}{3}\right) \times 0.3$$

$$= \frac{1}{3} \times \frac{2}{3} \times 0.3 = 0.07$$

The instructor is more likely to be in the office.