6/7/23

Assignment-2.

Elsamaria
SGAD
U2008029

2. Chest pain.

Yes : Y = 3 N = 0 T = 3

No : Y = 1 N = 2 T = 3

$E(Yes) = -\left[\frac{3}{3}\log(1) + 0\right] = 0$

$E(No) = -\left[\frac{1}{3}\log(1/3) + \frac{2}{3}\log(2/3)\right] = 0.918$

$E(parent) = -\left[\frac{4}{6}\log(4/6) + \frac{2}{6}\log(2/6)\right] = 0.9182$

$IG = 0.9182 - \left[\frac{3}{6}\times 0 + \frac{3}{6}\times 0.9182\right] = 0.4591$

Male

Yes : Y = 2 N = 2 T = 4

No : Y = 2 N = 0 T = 2

$E(Yes) = -\left[\frac{2}{4}\log(2/4) + \frac{2}{4}\log(2/4)\right] = 1$

$E(No) = -\left[\frac{2}{2}\log(1) + 0\right] = 0$

$IG = 0.9182 - \left[\frac{4}{6}\times 1 + \frac{2}{6}\times 0\right] = 0.2515$

Smokes

Yes : Y = 3 N = 1 T = 4

No : Y = 1 N = 1 T = 2

$E(Yes) = -\left[\frac{3}{4}\log(3/4) + \frac{1}{4}\log(1/4)\right] = 0.8112$

$E(No) = -\left(\frac{1}{2}\log(1/2) + \frac{1}{2}\log(1/2)\right) = 1$

$IG = 0.044406$

Exercise

Yes : Y = 2 N = 2 T = 4

No : Y = 2 N = 0 T = 2

$E(Yes) = -\left[\frac{2}{4}\log(2/4) + \frac{2}{4}\log(2/4)\right] = 1$

$E(No) = -\left(\frac{2}{2}\log(1)\right) = 0$

$IG = 0.2515$

⇒ Chest pain has highest I Gr.

∴ Yes

PID	Male	Exercise	Heart Smoke	Heart attack
1	Y	Y	N	Y
2	Y	N	Y	Y
5	N	Y	Y	Y

No

PID	Male	Exercise	Smoke	Heart attack.
3	N	N	Y	Y
4	Y	Y	N	N
6	Y	Y	Y	N

E(Male

$$E(Yes) = -\left[\frac{0}{2} + \frac{2}{2}\log(1)\right] = 0$$

$$E(No) = -\left[1\log(1) + 0\right] = 0$$

$$IG = 0.918$$

Date: / /

Exercise

$E(Yes) = -[0 + \frac{2}{2}\log(1)] = 0$

$E(No) = -[1\log(1) + 0] = 0$

$IG = 0.918$

Smoke

$E(Yes) = -[\frac{1}{2}\log(\frac{1}{2}) + \frac{1}{2}\log(\frac{1}{2})] = 1$

$E(No) = 0$

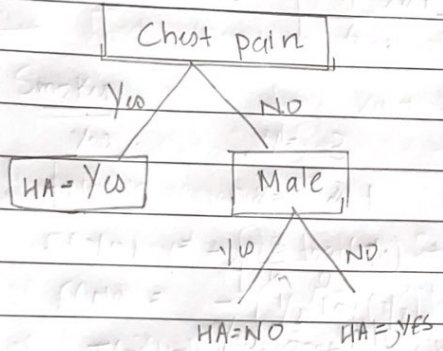
$IG = 0.2516$

⇒ Male & Exercise has highest IG

Male

Yes	PID	Exercise	Smoke	Heart attack
	4	Y	N	N
	6	Y	Y	N
No	PID	Exercise	Smoke	Heart attack
	3	N	Y	Y

Decision tree



1. $n = 400$

$300 \rightarrow 0$

$400 \rightarrow 1$

$$\text{Entropy} = -\sum p_i \log_2 p_i$$

$$= -\left[\frac{100}{400}\log\left(\frac{100}{400}\right) + \frac{300}{400}\log\left(\frac{300}{400}\right)\right]$$

$$= 0.81125$$

3. $$\text{Gain ratio} = \frac{IG}{\text{Intrinsic information}}$$

$$\text{Intrinsic info} = -\left[\frac{100}{200}\log\left(\frac{100}{200}\right) + \frac{100}{200}\log\left(\frac{100}{200}\right)\right]$$

$$= 1$$

A

$$\text{Gain ratio} = \frac{0.6}{1} = 0.6$$

B

$$\text{Gain ratio} = 0.4$$

C Gain ratio = 0.8

∴ Feature C has gain ratio.

4. $$E(\text{parent}) = -\left[\frac{3}{10}\log(3/10) + \frac{3}{10}\log(3/10) + \frac{4}{10}\log(4/10)\right]$$

$$= 1.5710$$

Sore throat

Yes

$$E(\text{Sore throat}) = -\left[\frac{2}{5}\log\left(\frac{2}{5}\right) + \frac{1}{5}\log\left(\frac{1}{5}\right) + \frac{2}{5}\log\left(\frac{2}{5}\right)\right]$$

$$= 1.5219$$

$$E(\text{No}) = 1.5219$$

$$IG = 1.57 - \left[1.5219 \times \frac{1}{2} + 1.5219 \times \frac{1}{2}\right] = 0.0481$$

Fever

$$E(Yes) = -\left[\frac{1}{4}\log\left(\frac{1}{4}\right) + 0 + \frac{3}{4}\log\left(\frac{3}{4}\right)\right] = 0.81127$$

$$E(No) = -\left[\frac{2}{6}\log\left(\frac{2}{6}\right) + \frac{3}{6}\log\left(\frac{3}{6}\right) + \frac{1}{6}\log\left(\frac{1}{6}\right)\right]$$

$$= 1.4591$$

$$IG = 0.57 - \left[0.81127 \times \frac{4}{10} + 1.4591 \times \frac{6}{10}\right] = 0.370032$$

Swollen Glands

$$E(Yes) = -\left[\frac{3}{3}\log\left(\frac{3}{3}\right) + 0 + 0\right] = 0$$

$$E(No) = -\left[0 + \frac{3}{7}\log\left(\frac{3}{7}\right) + \frac{4}{7}\log\left(\frac{4}{7}\right)\right] = 0.98522$$

$$IG = 1.57 - \left[0.98522 \times \frac{7}{10}\right] = 0.880346$$

Congestion

$$E(Yes) = -\left[\frac{1}{8}\log\left(\frac{1}{8}\right) + \frac{3}{8}\left(\log\left(\frac{3}{8}\right) + \frac{4}{8}\log\left(\frac{4}{8}\right)\right]\right.$$

$$= 1.4056$$

$$E(No) = -\left[\frac{2}{2}\log\left(\frac{2}{2}\right)\right] = 0$$

$$IG = 1.57 - \left[1.4056 \times \frac{8}{10}\right] = 0.44552$$

Headache

$$E(Yes) = -\left[\frac{1}{5}\log\left(\frac{1}{5}\right) + \frac{2}{5}\log\left(\frac{2}{5}\right) + \frac{2}{5}\log\left(\frac{2}{5}\right)\right)$$

$$= 1.5219$$

$$E(No) = -\left[\frac{2}{5}\log\left(\frac{2}{5}\right) + \frac{1}{5}\log\left(\frac{2}{5}\right) + \frac{2}{5}\log\left(\frac{2}{5}\right)\right]$$

$$= 1.5219$$

$$IG = 1.57 - \left[1.5219 \times \frac{1}{\alpha} + 1.5219 \times \frac{1}{\alpha}\right] = 0.04181$$

papergrid

Date: / /

⇒ Swollen Glands has the highest IG.

Yes

Sore throat	Fever	Congestion	Headache	Diagnosis
Y	Y	Y	Y	Strep throat
N	N	N	N	Strep throat
Y	N	N	N	Strep throat

No

Sore throat	Fever	Congestion	Headache	Diagnosis
N	N	Y	Y	Allergy
N	Y	Y	N	Cold
N	N	Y	N	Allergy
Y	N	Y	Y	Allergy
N	Y	Y	Y	Cold
Y	N	Y	Y	Cold
Y	Y	Y	N	Cold.

For Sore throat

$$E(Yes) = -\left[0 + \frac{1}{3}\log\left(\frac{1}{3}\right) + \frac{2}{3}\log\left(\frac{2}{3}\right)\right] = 0.9182$$

$$E(No) = -\left[0 + \frac{2}{4}\log\left(\frac{2}{4}\right) + \frac{2}{4}\log\left(\frac{2}{4}\right)\right] = 1$$

$$IG = 1.57 - \left[\frac{2}{10} + \frac{4}{10}\right] = 0.02$$

Fever.

$$E(Yes) = -\left[0 + 0 + \frac{3}{3}\log\left(\frac{3}{3}\right)\right] = 0$$

$$E(No) = -\left[0 + \frac{3}{4}\log\left(\frac{3}{4}\right) + \frac{1}{4}\log\left(\frac{1}{4}\right)\right] = 0.81127$$

$$IG = 0.52163$$

Congestion

$$E(Yes) = -\left[0 + \frac{3}{7}\log\left(\frac{3}{7}\right) + \frac{4}{7}\log\left(\frac{4}{7}\right)\right] = 0.985$$

$$E(No) = 0$$

$$IG = 0.98522 - 0.985 = 0$$

Headache.

$$E(Yes) = -\left[0 + \frac{2}{4}\log\left(\frac{2}{4}\right) + \frac{2}{4}\log\left(\frac{2}{4}\right)\right] = 1$$

$$E(No) = -\left[0 + \frac{1}{3}\log\left(\frac{1}{3}\right) + \frac{2}{3}\log\left(\frac{2}{3}\right)\right] = 0.9182$$

~~IG = 0.005~~

$IG = 0.02027$

⇒ Fever has the highest IG.

Yes.

Sore throat	Congestion	Headache	Diagonsis.
N	Y	N	Cold
N	Y	Y	Cold
Y	Y	N	Cold

No

Sore throat	Congestion	Headache	Diagonsis
N	Y	Y	Allergy
N	Y	N	Allergy
Y	Y	Y	Allergy
Y	Y	Y	Cold.

E(parent) = 0.81127

Sore throat

$$E(Yes) = -\left[0 + \frac{1}{2}\log(1/2) + 1/2\log(1/2)\right] = 1$$

$$E(No) = -\left[0 + \frac{2}{2}\log(1/2)\right] = 0$$

$$IG = 0.81127 - \frac{2}{4} = 0.31127$$

Congestion

$$E(Yes) = -\left[0 + 3/4\log(3/4) + 1/4\log(1/4)\right] = 0.81127$$

$E(No) = 0$

$IG = 0.81127 - 0.81127 = 0$

~~IG = 0.51127~~

Headache.

$$E(Yes) = -\left[\frac{2}{3}\log(2/3) + 1/3\log(1/3)\right] = 0.91829$$

$$E(No) = [1\log(1)] = 0$$

$$IG = 0.1225$$

→ Sore throat has the highst IG.

Yes	Congestion	Headache	Diagonsis
	Y	Y	Allugy
	Y	Y	cold

No	Congestion	Headache	Diagonsis
	Y	Y	Allugy
	Y	N	Allugy

Yes ⇒ $E(parent) = -\left(\frac{1}{2}\log(1/2) + 1/2\log(1/2)\right] = 1$

Congestion

$E(Yes) = 1$

$E(No) = 0$

$IG = 1 - 1 = 0$

Headache

$E(Yes) = 1$

$E(No) = 0$

$IG = 0$