Using chi-squared to make a decision tree

Step 1)

Compute chi-squared of each attribute, highest being root node.

Headache

Yes = 4 | No = 6

P=3, N=1 | P=2, N=4

Since there is a 50/50 split in Diagnostics, expected is 0.5 x total

Headache	Positive	Negative	Total
Yes	3 (0.5 x 4 = 2)	1 (0.5 x 4 = 2)	4
No	2 (0.5 x 6 = 3)	4 (0.5 x 6 = 3)	6
Total	5	5	10

$$\frac{(3-2)^2}{2} + \frac{(1-2)^2}{2} + \frac{(2-3)^2}{3} + \frac{(4-3)^2}{3}$$

$$= 5/3 = 1.666\ldots$$

Spots

Yes = 6 | No = 4

P=4, N=2 | P=1, N=3

Spots	Positive	Negative	Total
Yes	4 (0.5 x 6 = 3)	2 (0.5 x 6 = 3)	6
No	1 (0.5 x 4 = 2)	3 (0.5 x 4 = 2)	4
Total	5	5	10

$$\frac{(4-3)^2}{3} + \frac{(2-3)^2}{3} + \frac{(1-2)^2}{2} + \frac{(3-2)^2}{2}$$

$$= 5/3 = 1.666\ldots$$

Stiff-neck

Yes = 5 (P=4, N=1)

No = 5 (P=1, N=4)

Stiff-neck	Positive	Negative	Total
Yes	4(0.5×5=2.5)	1(0.5×5=0.25)	5
No	1(0.5×5=2.5)	4(0.5×5=0.25)	5
Total	5	5	

$$\frac{(4-2.5)^2}{2.5} + \frac{(1-2.5)^2}{2.5} + \frac{(1-2.5)^2}{2.5} + \frac{(4-2.5)^2}{2.5}$$

$$= 18/5 = 3.6$$

Headache = 1.667 (3.s.f)

Spots = 1.667 (3.s.f)

Stiff-Neck = 3.600 (3.s.f) ← root node.

Stiff-neck

Yes

Cid	H	Sp	S	D
1	Y	Y	Y	P
3	Y	N	Y	P
4	N	Y	Y	P
5	Y	Y	Y	P
6	N	N	Y	N

No

Cid	H	Sp	S	D
2	N	Y	N	P
7	N	Y	N	N
8	N	Y	N	N
9	Y	N	N	N
10	N	N	N	N

Stiff neck = Yes

P=4, N=1 ∴ 4:1

$\frac{4}{5}$ = True , $\frac{1}{5}$ = false.

Headache

Yes = 3 N = 2

P = 3, N = 0 P = 1, N = 1

Headache	Positive	Negative	Total
Yes	3 ($4/5 \times 3 = 12/5$)	0 ($1/5 \times 3 = 3/5$)	3
NO	1 ($4/5 \times 2 = 8/5$)	1 ($1/5 \times 2 = 2/5$)	2
Total	4	1	5

$$\frac{(3 - \frac{12}{5})^2}{(\frac{12}{5})} + \frac{(0 - 3/5)^2}{3/5} + \frac{(1 - 8/5)^2}{8/5} + \frac{(1 - 2/5)^2}{2/5}$$

$= 15/8 = 1.875$

Spot 7

Yes = 3 NO = 2

P = 3, N = 0 P = 1, N = 1

Spot	Positive	Negative	Total
Yes	3 ($4/5 \times 3 = 12/5$)	0 ($1/5 \times 3 = 3/5$)	3
NO	1 ($4/5 \times 2 = 8/5$)	1 ($1/5 \times 2 = 2/5$)	2
Total	4	1	5

$$\frac{(3 - 12/5)^2}{12/5} + \frac{(0 - 3/5)^2}{3/5} + \frac{(1 - 8/5)^2}{8/5} + \frac{(1 - 2/5)^2}{2/5}$$

$= 15/8 = 1.875$

∴ Both matter equally.

Stiff-neck = No

P = 1, N = 4 ∴ 1:4

1/5 = True, 4/5 = False.

Headache

Yes = 1 | No = 4

~~P = 1~~, N = 1 | P = 1, N = 3

Headache	Positive	Negative	Total
Yes	0 (1/5 × 1 = 1/5)	1 (4/5 × 1 = 4/5)	1
No	1 (1/5 × 4 = 4/5)	3 (4/5 × 4 = 16/5)	4
Total	1	4	5

$$\frac{(0-1/5)^2}{1/5} + \frac{(1-4/5)^2}{4/5} + \frac{(1-4/5)^2}{4/5} + \frac{(3-16/5)^2}{16/5}$$

$$= \frac{2.1}{10} = 2.1 = 5/16 = 0.3125$$

Spots

Y = 3 | N = 2

P = 1, N = 2 | P = 0, N = 2

SPots	Positive	Negative	Total
Yes	1 (1/5 × 3 = 3/5)	2 (4/5 × 3 = 12/5)	3
No	0 (1/5 × 2 = 2/5)	2 (4/5 × 2 = 8/5)	2
Total	1	4	5

$$\frac{(1-3/5)^2}{3/5} + \frac{(2-12/5)^2}{12/5} + \frac{(0-2/5)^2}{2/5} + \frac{(2-8/5)^2}{8/5}$$

$$= 5/6$$

$$= 0.833\ldots$$

~~so split on Headache.~~ So split on Spots.

Spots

Yes

Cid	H	Sp	S	D
2	N	Y	N	P
7	N	Y	N	N
8	N	Y	N	N

↑

2/3 chance of being negative.

No

Cid	H	Sp	S	D
9	Y	N	N	N
10	N	N	N	N

↑ stop

Chi-square Decision tree