① a)

F1	F2	Color	Result
L	H	R	good
L	L	B	bad
H	H	B	good
L	H	B	bad
H	L	R	good

b) $S = [3^+, 2^-]$; $p^+ = .6$, $p^- = .4$

$E(S) = -.6 \log_2 .6 - .4 \log_2 .4$

≈ 0.971

$\boxed{S_{F1}}$

$S_{F1=H} = [2^+, 0^-]$

$S_{F1=L} = [1^+, 2^-]$

$E(S_{F1=H}) = -\log_2 1 = 0$

$E(S_{F1=L}) = -\frac{1}{3}\log_2 \frac{1}{3} - \frac{2}{3}\log_2 \frac{2}{3} = .719$

$$IG(S, F1) = E(S) - \left(\frac{S_{F1=H}}{S}\right) E(S_{F1=H}) - \left(\frac{S_{F1=L}}{S}\right) E(S_{F1=L})$$

$= .971 - 0 - (3/5).719$

$= .540$

$\boxed{S_{F2}}$

$$S_{F2=H} = [2^+, 1^-]$$

$$S_{F2=L} = [1^+, 1^-]$$

$$E(S_{F2=H}) = -\tfrac{2}{3}\log_2 \tfrac{2}{3} - \tfrac{1}{3}\log_2 \tfrac{1}{3} = .719$$

$$E(S_{F2=L}) = 2\left(-\tfrac{1}{2}\log_2 \tfrac{1}{2}\right) = .89\phi$$

$$IG(S, F2) = .971 - (\tfrac{3}{5}).719 - (\tfrac{2}{5}).89\phi$$

$$= .184$$

$\boxed{S_{color}}$

$$S_{Color=R} = [2^+, \phi^-]$$

$$S_{Color=B} = [1^+, 2^-]$$

$$S_{Color=Y} = [\phi^+, \phi^-]$$

$$E(S_{Color=R}) = -1\log_2 1 = \phi$$

$$E(S_{Color=B}) = -\tfrac{1}{3}\log_2 \tfrac{1}{3} - \tfrac{2}{3}\log_2 \tfrac{2}{3} = .719$$

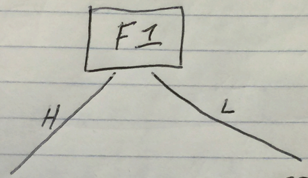
$$E(S_{Color=Y}) = \phi$$

$$IG(S, Color) = .971 - (\tfrac{2}{5})\phi - (\tfrac{3}{5}).719 - (\tfrac{\phi}{5})\phi$$

$$= .54\phi$$

Feature	IG
F1	.540
F2	.184
Color	.540

Root node is F1:
- Highest IG Value
- Tie broken over Color



F2	Color	Result
H	B	good
L	R	good

F2	Color	Result
H	R	good
L	B	bad
H	B	bad

~~Entro~~

$$\cancel{S_{[F1=H,\,F2=H]} = S_{[F1=H,\,F2=L]} = [1^+, 0^-]}$$

$$\cancel{E(S_{[F1=H,\,F2=H]}) = E(S_{[F1=H,\,F2=L]}) = -1\log_2 1 = 0}$$

$$\cancel{IG(S,[F1=H,F2]) = .971 - 0 - 0 = .971}$$

Yes

next page

$E(S_{F1=L}) = -\frac{1}{3}\log_2 \frac{1}{3} - \frac{2}{3}\log_2 \frac{2}{3} = .7\ldots$

$S_{F1=L, F2=H} = [1^+, 1^-]$

$S_{F1=L, F2=L} = [0^+, 1^-]$

$E(S_{F1=L, F2=H}) = 2\left(-\frac{1}{2}\log_2 \frac{1}{2}\right) = .890$

$E(S_{F1=L, F2=L}) = -1 \log 1 = 0$

$IG(S_{F1=L}, F2) = \underbrace{.719}_{E(S_{F1=L})} - \cancel{.890}\ \frac{2}{3}\underbrace{(.890)}_{E(S_{F1=L},F2=H)} - 0$

$= .126$

$S_{F1=L, Color=R} = [1^+, 0^-]$

$S_{F1=L, Color=B} = [0^+, 2^-]$

$S_{F1=L, Color=Y} = [0^+, 0^-]$

$E(S_{F1=L, Color=R}) = 0$ $(= E(S_{F1=L, F2=L}))$

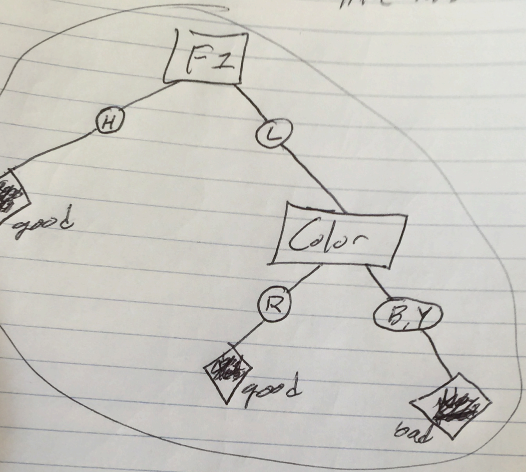
$E(S_{F1=L, Color=B}) = 0$ $(= E(S_{F1=H}))$

$E(S_{F1=L, Color=Y}) = 0$ (nothing to choose from)

$IG(S_{F1=L}, Color) = .719 - 0 - 0 - 0$

$= .719$

Highest is Color, use that as next ~~att~~ node o, final tree looks like this:



The prediction accuracy can be determined by running it through the tree and calculating the number of correct results:

F1(H/L)	F2	F2(H/L)	Color	Result	Tree Result
L	4	H	R	good	good
L	0	L	B	bad	bad
H	7	H	Y	good	good

ecause all results match the results from the tree, the prediction accuracy can be determined to be 100%.