

GAIN RATIO

Height	Hair	Eyes	Class
Short	Blond	Blue	+
Tall	Blond	Brown	-
Tall	Red	Blue	+
Short	Dark	Blue	-
Tall	Dark	Blue	-
Tall	Blond	Blue	+
Tall	Dark	Brown	-
Short	Blond	Brown	-

$$\text{Info}(D) = I(3,5)$$

$$= -\frac{3}{8} \log_2 \left(\frac{3}{8}\right) - \frac{5}{8} \log_2 \left(\frac{5}{8}\right)$$

$$= 0.954$$

$$\begin{aligned}
 \text{Info}_{\text{height}}(D) &= \frac{3}{8} I(1,2) + \frac{5}{8} I(2,3) \\
 &= \frac{3}{8} \left(-\frac{1}{3} \log\left(\frac{1}{3}\right) - \frac{2}{3} \log\left(\frac{2}{3}\right) \right) + \\
 &\quad \frac{5}{8} \left(-\frac{2}{5} \log\left(\frac{2}{5}\right) - \frac{3}{5} \log\left(\frac{3}{5}\right) \right) \\
 &= 0.951
 \end{aligned}$$

$$\begin{aligned}
 \text{Info}_{\text{hair}}(D) &= \frac{4}{8} I(2,2) + \frac{1}{8} I(1,0) + \frac{3}{8} I(0,3) \\
 &= \frac{4}{8} \left(-\frac{2}{4} \log\left(\frac{2}{4}\right) - \frac{2}{4} \log\left(\frac{2}{4}\right) \right) + \\
 &\quad \frac{1}{8} \left(-\frac{1}{1} \log\left(\frac{1}{1}\right) - \frac{0}{1} \log\left(\frac{0}{1}\right) \right) + \\
 &\quad \frac{3}{8} \left(-\frac{0}{3} \log\left(\frac{0}{3}\right) - \frac{3}{3} \log\left(\frac{3}{3}\right) \right) \\
 &= 0.5
 \end{aligned}$$

$$\begin{aligned}
 \text{Info}_{\text{eyes}}(D) &= \frac{5}{8} I(3,2) + \frac{3}{8} I(0,3) \\
 &= \frac{5}{8} \left(-\frac{3}{5} \log\left(\frac{3}{5}\right) - \frac{2}{5} \log\left(\frac{2}{5}\right) \right) + \\
 &\quad \frac{3}{8} \left(-0 - \frac{3}{3} \log\left(\frac{3}{3}\right) \right) \\
 &= 0.606
 \end{aligned}$$

$$\begin{aligned} \text{Gain}_{\text{height}} &= \text{Info}(D) - \text{Info}_{\text{height}}(D) \\ &= 0.954 - 0.951 \\ &= \boxed{0.003} \end{aligned}$$

$$\begin{aligned} \text{Gain}_{\text{hair}} &= 0.954 - 0.5 \\ &= \boxed{0.454} \end{aligned}$$

$$\begin{aligned} \text{Gain}_{\text{eyes}} &= 0.954 - 0.606 \\ &= \boxed{0.348} \end{aligned}$$

Now find split info

$$\begin{aligned} \text{Splitinfo}_{\text{height}}(D) &= -\frac{3}{8} \log\left(\frac{3}{8}\right) - \frac{5}{8} \log\left(\frac{5}{8}\right) \\ &= 0.954 \end{aligned}$$

$$\begin{aligned} \text{Splitinfo}_{\text{hair}}(D) &= -\frac{4}{8} \log\left(\frac{4}{8}\right) - \frac{1}{8} \log\left(\frac{1}{8}\right) - \\ &\quad \frac{3}{8} \log\left(\frac{3}{8}\right) \\ &= 1.405 \end{aligned}$$

$$\text{SplitInfo}_{\text{(eye)}} = -\frac{5}{8} \log\left(\frac{5}{8}\right) - \frac{3}{8} \log\left(\frac{3}{8}\right)$$

$$= 0.954$$

Now find GuinRatio :

$$\text{GuinRatio}_{\text{(height)}} = \frac{\text{Guin}_{\text{(height)}}}{\text{SplitInfo}_{\text{(height)}}}$$

$$= \frac{0.003}{0.954}$$

$$= 0.00314$$

$$\text{GuinRatio}_{\text{(hair)}} = \frac{0.454}{1.405}$$

$$= 0.3231$$

$$\text{GuinRatio}_{\text{(eye)}} = \frac{0.348}{0.954}$$

$$= 0.3647$$

Highest GuinRatio is selected as a root node.

EYES

Height	Hair	class
Short	Blond	+
Tall	Red	+
Short	Dark	-
Tall	Dark	-
Tall	Blond	+

Height	Hair	class
Tall	Blond	-
Tall	Dark	-
Short	Blond	-

$$\text{Info}(D) = 0.97$$

Now again apply same procedure for Blue.

$$\begin{aligned}
 \text{Info}_{(\text{height})}(D) &= \frac{2}{5} I(1,1) + \frac{3}{5} I(2,1) \\
 &= \frac{2}{5} \left(-\frac{1}{2} \log\left(\frac{1}{2}\right) - \frac{1}{2} \log\left(\frac{1}{2}\right) \right) + \\
 &\quad \frac{3}{5} \left(-\frac{2}{3} \log\left(\frac{2}{3}\right) - \frac{1}{3} \log\left(\frac{1}{3}\right) \right) \\
 &= 0.95
 \end{aligned}$$

$$\begin{aligned}
 \text{Info (D)}_{\text{(hair)}} &= \frac{2}{5} I(2,0) + \frac{1}{5} I(1,0) + \frac{2}{5} I(0,2) \\
 &= \frac{2}{5} \left(-\frac{2}{2} \log\left(\frac{2}{2}\right) \cdot -0 \right) + \\
 &\quad \frac{1}{5} \cdot \dots \cdot 0 + 0 \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 \text{Gini}_{\text{(height)}} &= 0.97 - 0.95 \\
 &= 0.02
 \end{aligned}$$

$$\begin{aligned}
 \text{Gini}_{\text{(hair)}} &= 0.97 - 0 \\
 &= 0.97
 \end{aligned}$$

Now find split Info:

$$\begin{aligned}
 \text{Split Info}_{\text{(height)}} &= -\frac{2}{5} \log\left(\frac{2}{5}\right) - \frac{3}{5} \log\left(\frac{3}{5}\right) \\
 &= 0.97
 \end{aligned}$$

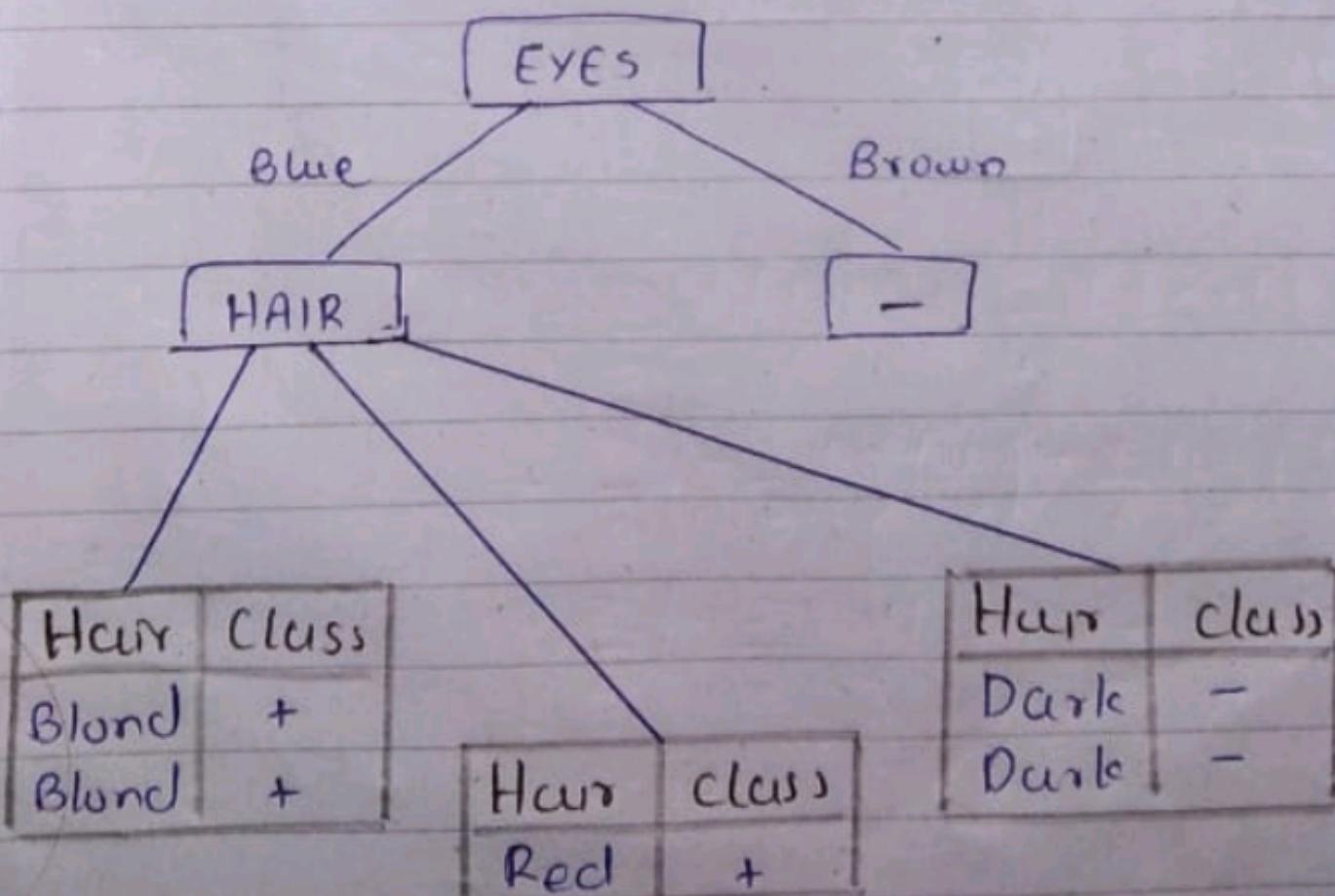
$$\begin{aligned}
 \text{Split Info}_{\text{(hair)}} &= -\frac{2}{5} \log\left(\frac{2}{5}\right) - \frac{1}{5} \log\left(\frac{1}{5}\right) - \frac{2}{5} \log\left(\frac{3}{5}\right) \\
 &= 1.15
 \end{aligned}$$

Now find Guin Ratio

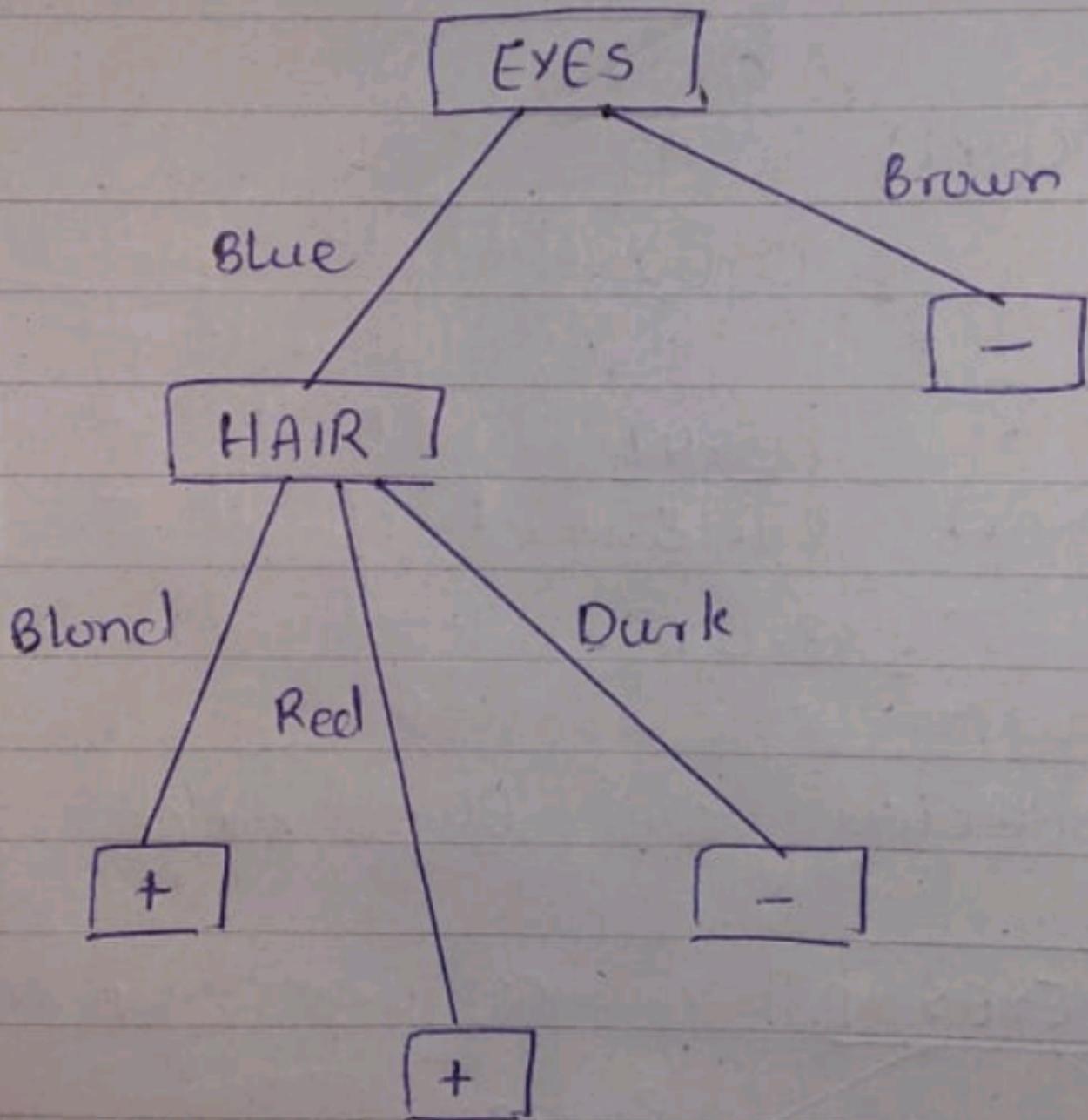
$$\text{GuinRatio}_{(\text{height})} = \frac{0.02}{0.97} = 0.02$$

$$\text{GuinRatio}_{(\text{hair})} = \frac{0.97}{1.15} = 0.843$$

hair is selected as blue nucle



Final Tree !



ASSIGNMENT

Date _____

NAME: MAHNOOR SHAKIL

CLASS: BSSE - III

SEAT NO: B18158024

QUESTION : 01

GINI INDEX

Attribute 1	Attribute 2	Attribute 3	Class
A	70	T	C1
A	90	T	C2
A	85	F	C2
A	95	F	C2
A	70	F	C1
B	90	T	C1
B	78	F	C1
B	65	T	C1
B	75	F	C1
C	80	T	C2
C	70	T	C2
C	80	F	C1
C	80	F	C1
C	96	F	C1

Binary splitting Attribute 1:

Attribute 1 has 3 values so possible splits are;

- 1 {A, B}, {C}
- 2 {A, C}, {B}
- 3 {B, C}, {A}

Date _____

$$\text{Gini}_{\{A, B\}, \{C\}} = \frac{9}{14} \left\{ 1 - \left(\frac{6}{9} \right)^2 - \left(\frac{3}{9} \right)^2 \right\} + \frac{5}{14} \left\{ 1 - \left(\frac{3}{5} \right)^2 - \left(\frac{2}{5} \right)^2 \right\}$$

$$= 0.457$$

$$\text{Gini}_{\{A, C\}, \{B\}} = \frac{10}{14} \left\{ 1 - \left(\frac{5}{10} \right)^2 - \left(\frac{5}{10} \right)^2 \right\} + \frac{4}{14} \left\{ 1 - \left(\frac{4}{4} \right)^2 - \left(\frac{0}{4} \right)^2 \right\}$$

$$= 0.357$$

$$\text{Gini}_{\{B, C\}, \{A\}} = \frac{9}{14} \left\{ 1 - \left(\frac{7}{9} \right)^2 - \left(\frac{2}{9} \right)^2 \right\} + \frac{5}{14} \left\{ 1 - \left(\frac{2}{5} \right)^2 - \left(\frac{3}{5} \right)^2 \right\}$$

$$= 0.393$$

Thus, amongst the three possible splits of Attribute 1, the gini of $\{A, C\}, \{B\}$ is the lowest.

Now, two possible outcomes of Attribute 1 are:

1. $\{A, C\}$
2. $\{B\}$

Binary Splitting Attribute 2:

Class C1 C1 C1 C1 C2 C1 C1 C2 C1 C1 C2 C2 C2 C2 C1

Sorted 65 70 70 70 75 78 80 80 80 85 90 90 95 96

Values

Split 60 67 70 70 72 76 79 80 80 82 87 90 92 95

Positions

$\leq > \leq >$

C1 0 9 1 8 3 6 3 6 3 6 4 5 5 4 7 2 7 2 7 2 7 2 8 1 8 1 8 1

C2 0 5 0 5 1 4 1 4 1 4 1 4 1 4 2 3 2 3 2 3 3 2 4 1 4 1 5 0

Gini 0.459 0.439 0.45 0.45 0.45 0.431 0.404 0.393 0.393 0.393 0.442 0.452 0.452 0.439

} After binary splitting table becomes;

Attribute 1	Attribute 2	Attribute 3	Class
{A, C}	≤ 80	T	C1
{A, C}	> 80	T	C2
{A, C}	> 80	F	C2
{A, C}	> 80	F	C2
{A, C}	≤ 80	F	C1
B	> 80	T	C1
B	≤ 80	F	C1
B	≤ 80	T	C1
B	≤ 80	F	C1
{A, C}	≤ 80	T	C2
{A, C}	≤ 80	T	C2
{A, C}	≤ 80	F	C1
{A, C}	≤ 80	F	C1
{A, C}	> 80	F	C1

$$\text{Gini attribute 3} = \frac{T}{14} \left\{ 1 - \left(\frac{3}{6} \right)^2 - \left(\frac{3}{6} \right)^2 \right\} + \frac{F}{14} \left\{ 1 - \left(\frac{6}{8} \right) - \left(\frac{2}{8} \right)^2 \right\}$$

$$= 0.428$$

$$\text{Gini class} = 1 - \left(\frac{9}{14} \right)^2 - \left(\frac{5}{14} \right)^2 = 1 - \left(\frac{81}{196} \right) - \left(\frac{25}{196} \right)$$

$$= 0.459$$

$$\text{Gini attribute 1} = 0.357$$

$$\text{Gini attribute 2} = 0.393$$

Since Gini attribute 1 has the smallest value so it will be chosen as the root node.

Date _____

attribute 1?

{A, C}

{B}

	Attribute 2	Class	Attribute 3
{A, C}	≤ 80	C1	T
{A, C}	> 80	C2	T
{A, C}	> 80	C2	F
{A, C}	> 80	C2	F
{A, C}	≤ 80	C1	F
{A, C}	≤ 80	C2	T
{A, C}	≤ 80	C2	T
{A, C}	≤ 80	C1	F
{A, C}	≤ 80	C1	F
{A, C}	> 80	C1	F

	Attribute 2	Attribute 3	Class
{B}	> 80	T	C1
{B}	≤ 80	F	C1
{B}	≤ 80	T	C1
{B}	≤ 80	F	C1

{A, C}:

$$\text{Gini attribute 2} = \frac{6}{10} \left\{ 1 - \left(\frac{1}{6} \right)^2 - \left(\frac{2}{6} \right)^2 \right\} + \frac{4}{10} \left\{ 1 - \left(\frac{1}{4} \right)^2 - \left(\frac{3}{4} \right)^2 \right\}$$

$$= 0.416$$

$$\text{Gini attribute 3} = \frac{4}{10} \left\{ 1 - \left(\frac{1}{4} \right)^2 - \left(\frac{3}{4} \right)^2 \right\} + \frac{6}{10} \left\{ 1 - \left(\frac{4}{6} \right)^2 - \left(\frac{2}{6} \right)^2 \right\}$$

$$= 0.416$$