Example

Dataset.

$$Gini(D) = 1 - \left(\left(\frac{9}{14}\right)^2 + \left(\frac{5}{14}\right)^2\right)$$

$$= 1 - [0.4132 + 0.1275)$$
$$= 0.4593.$$

Income

$D_1 = \{$ low, medium $\}$ & $\{$ high $\}$.

$$Gini(D_1) = \frac{10}{14} Gini_{(Low, medium)} + \frac{4}{14} Gini(High)$$

$$= \frac{10}{14}\left(1 - \left(\left(\frac{1}{10}\right)^2 + \left(\frac{3}{10}\right)^2\right)\right) + \frac{4}{14}$$

$$\left(1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2\right)$$

$$= \frac{10}{14}\left(1 - 0.44 - 0.09\right) + \frac{4}{14}(1 - 0.5)$$

$$= 0.3 + 0.142$$
$$= 0.4420.$$

→ D_2 = {low, high} & {medium}.

$$Gini(D_2) = \frac{8}{14} Gini_{(low,high)} + \frac{6}{14} Gini_{(medium)}$$

$$= \frac{8}{14}\left[1 - \left(\left(\frac{5}{8}\right)^2 + \left(\frac{3}{8}\right)^2\right)\right] + \frac{6}{14}\left[1 - \left(\frac{2}{6}\right)^2 - \left(\frac{4}{6}\right)^2\right]$$

$$= \frac{8}{14}\left[1 - 0.5\right] + \frac{6}{14}\left[1 - 0.1111 - 0.4444\right]$$

$$= 0.026785 + 0.19057$$

$$= 0.45852.$$

→ D_3 = {medium, high} & {low}.

$$Gini(D_3) = \frac{10}{14}\left[1 - \left(\left(\frac{6}{10}\right)^2 + \left(\frac{4}{10}\right)^2\right)\right] + \frac{4}{14}\left[1 - \left(\left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2\right)\right]$$

$$= \frac{10}{14}\left[1 - 0.36 - 0.16\right] + \frac{4}{14}\left[1 - 0.0625 - 0.5625\right)$$

$$= 0.34285 + 0.1071$$

$$= 0.44995$$

$$\Delta Gini(Income) = 0.4593 - 0.4420$$

$$= 0.0173$$

Age

→ $D_1 = \{youth, Adult\}$ & $\{senior\}$.

$$Gini(D_1) = \frac{9}{14} Gini_{(Y,A)} + \frac{5}{14} Gini(S)$$

$$= \frac{9}{14}\left[1 - \left(\left(\frac{6}{9}\right)^2 + \left(\frac{3}{9}\right)^2\right)\right] + \frac{5}{14}\left[1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2\right]$$

$$= \frac{9}{14}\left[1 - 0.5555\right] + \frac{5}{14}\left[1 - 0.52\right]$$

$$= 0.28575 + 0.17142$$

$$= 0.45717.$$

→ $D_2 = \{youth, Senior\}$ & $\{Adult\}$.

$$Gini(D_2) = \frac{10}{14} Gini_{(4,5)} + \frac{4}{14} Gini(A).$$

$$= \frac{10}{14}\left[1 - \left(\frac{5}{10}\right)^2 - \left(\frac{5}{10}\right)^2\right] + \frac{4}{14}\left[1 - \left(\frac{4}{4}\right)^2\right].$$

$$= \frac{10}{14}\left[1 - 0.5\right] + \frac{4}{14}\left[1 - 1\right]$$

$$= 0.35714$$

→ $D_3 = \{Senior, Adult\}$ & $\{youth\}$.

$$Gini(D_3) = \frac{9}{14} Gini(S,A) + \frac{5}{14} Gini(4).$$

$$= \frac{9}{14}\left[1 - \left(\frac{1}{9}\right)^2 - \left(\frac{2}{9}\right)^2\right] + \frac{5}{14}\left[1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2\right]$$

$$= \frac{9}{14}[0.3456] + \frac{5}{14}[0.48]$$

$$= 0.3936.$$

$$\Delta \text{Gini(Age)} = 0.4593 - 0.3514$$
$$= 0.10216$$

Student

$D_1 = \{\text{No}\}$ & $\{\text{Yes}\}$.

$$\text{Gini}(D_1) = \frac{7}{14}\text{Gini}(N) + \frac{7}{14}\text{Gini}(Y)$$

$$= \frac{7}{14}\left[1 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2\right] + \frac{7}{14}\left[1 - \left(\frac{1}{7}\right)^2 - \left(\frac{6}{7}\right)^2\right]$$

$$= \frac{7}{14}[0.48979] + \frac{7}{14}[0.24489]$$

$$= 0.3673.$$

$$\therefore \Delta\text{Gini(Student)} = 0.4593 - 0.3673$$
$$= 0.0919.$$

* Credit

$D_1 = \{ \text{fair} \}$ $P\{\text{Excellent}\}$

$$Gini(D_1) = \frac{8}{14} GiniF + \frac{6}{14} GiniE$$

$$= \frac{8}{14}\left[1 - \left(\frac{2}{8}\right)^2 - \left(\frac{6}{8}\right)^2\right] + \frac{6}{14}\left[1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2\right]$$

$$= 0.21428 + 0.21428$$

$$= 0.4287$$

$$\Delta Gini(credit) = 0.4593 - 0.4287$$

$$= 0.03072.$$

→ ~~Age provides maximum reduction in impurity so Age is root Node~~

Attribute	Split	Gini Index
Income	{high}	0.449
	{mid, low}	
Age	{adult}	0.343
Student	binary	0.367
Credit	binary	0.428

→ Age has maximum reduction in impurity.

* for adult

Gini (D) = I.

→ Senior

$$Gini(D) = \left(1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2\right) = 0.48.$$

Income

$$Gini(D) = \frac{2}{5}\left(1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right) + \frac{3}{5}\left(1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2\right)$$

$$= 0.466.$$

→ Gini(medium)

→ Student

$$Gini(yes) = \frac{3}{5}\left(1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2\right) + \frac{2}{5}\left(1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right)$$

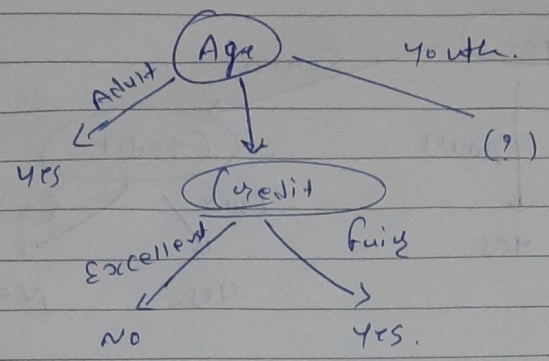
$$= 0.466.$$

→ Credit

$$Gini(fair) = \frac{5}{5}\left(1 - \left(\frac{3}{5}\right)^2\right) + \frac{2}{5}\left(1 - \left(\frac{1}{2}\right)^2\right)$$

$$= 0.$$

Credit has highest reduction purity.



→ for Youth

Income

$$\text{Gini(low)} = \frac{1}{5}\left(1-\left(\frac{1}{1}\right)^2-\left(\frac{0}{1}\right)^2\right) + \frac{4}{5}\left(1-\left(\frac{1}{4}\right)^2-\left(\frac{3}{4}\right)^2\right)$$

$$= 0.3$$

$$\text{Gini (medium)} = \frac{2}{5}\left(1-\left(\frac{1}{2}\right)^2-\left(\frac{1}{2}\right)^2\right) + \frac{3}{5}\left(1-\left(\frac{1}{3}\right)^2-\left(\frac{1}{3}\right)^2\right)$$

$$= 0.466.$$

$$\text{Gini (high)} = \frac{2}{5}\left(1-\left(\frac{2}{2}\right)^2\right) + \frac{3}{5}\left(1-\left(\frac{2}{3}\right)^2-\left(\frac{1}{3}\right)^2\right)$$

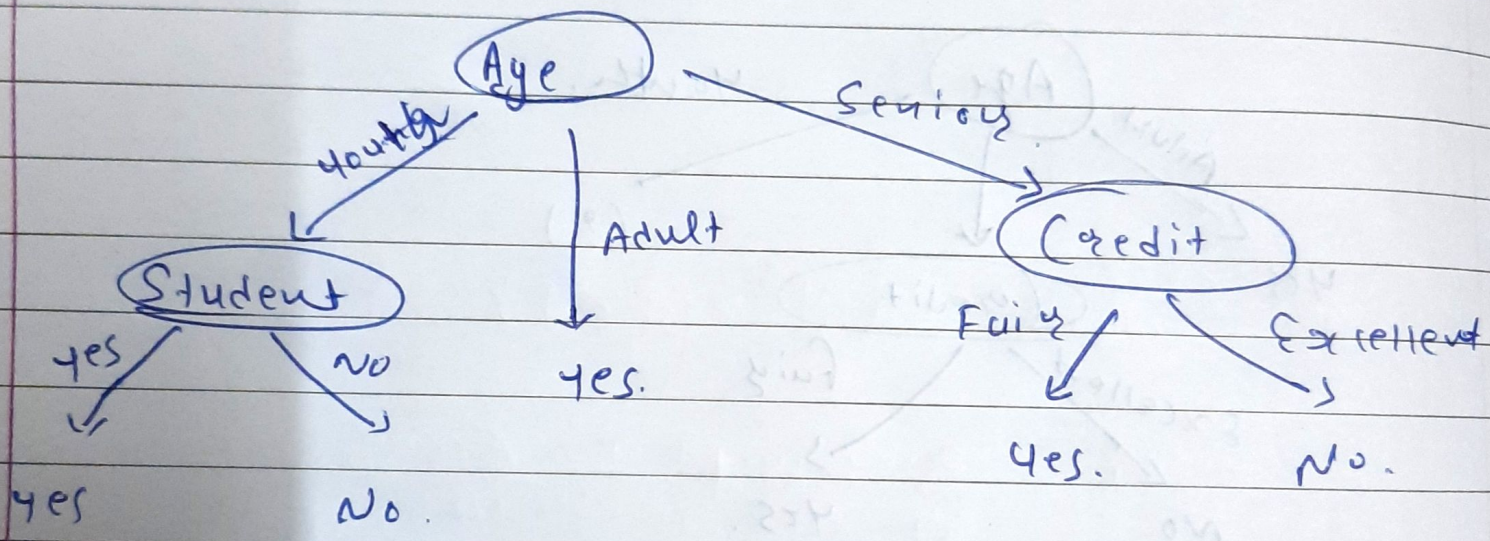
$$= 0.2666.$$

→ Student

$$\text{Gini (no)} = \frac{3}{5}\left(1-\left(\frac{3}{3}\right)^2\right) + \frac{2}{5}\left(1-\left(\frac{2}{2}\right)^2\right).$$

$$= 0.$$

→ Student has highest reduction in purity.


Age
Youth
Adult
Senior
Student
yes
No
yes
No.
yes.
Credit
Fair
Excellent
yes.
No.