Decision Tree using Gini Index.

Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay In
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

Allribute having minimum Gini Index is the attribute having maximum info gain

3 allributes - ~~weekend~~, weather, parents, Money.
Target variable - Decision
4 Possible ops - cinema
Tennis
Stay In
Shopping

We have to built Decision tree
~~select best~~
we need to calculate gini index of every attribute
allebute having min gini index is the attribute having max info gain

step 1: Calculate Gini of whole dataset

The data has 6 instances of Cinema
2 " of Tennis
1 " of Stay in
1 " " shopping

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

$$= \underline{0.58}$$

step 2: Calculate Gini index for every attribute

Attribute = Money

Value (Money) = Poor, Rich
Rich – 7 examples ~~8 cinema~~
Poor – 3 examples. 3 cinema

Money = Poor

$$Gini(S) = 1 - \left[\left(\frac{3}{3}\right)^2\right] = \underline{0}$$

Money = Rich

~~Gini~~ 3 instances are of cinema 2 Tennis
1 stay in
1 shopping

$$Gini(S)\ 1 - \left[\left(\frac{3}{7}\right)^2 + \left(\frac{2}{7}\right)^2 + \left(\frac{1}{7}\right)^2 + \left(\frac{1}{7}\right)^2\right]$$

$$= \underline{0.694}$$

Page No.

Weighted Average (Money)

$$= 0 * \left(\frac{3}{10}\right) + 0.694 \times \left(\frac{7}{10}\right) = 0.486$$

Allribute = Parents.

values = yes, no.

Parents = yes

5 examples all with cinema

$$\text{Gini}(S) = 1 - \left[\left(\frac{5}{5}\right)^2\right] = 0.$$

Parents = no.

2 example with tennis

1 " " stayon

1 " " shopping

1 " " cinema

$$\text{Gini}(S) = 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2\right]$$

$$= 0.72.$$

Weighted Average (Parents)

$$= 0 \times \left(\frac{5}{10}\right) + 0.72 \times \left(\frac{5}{10}\right) = 0.36$$

Date

Allubute = Weather

Weather = Sunny 2 examples with cinema
1 " " tennis

Gini (sunny) = $1 - \left[\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right] = 0.444$

Weather = Rainy 2 examples with cinema
1 " " stay in

Gini (Rainy) = $1 - \left[\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right] = 0.444$

Weather = Windy 3 examples with Cinema
1 " " Shopping

Gini (Windy) = $1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right] = 0.375$

Weighted Average (Weather)

$= 0.444 \times \frac{3}{10} + 0.444 \times \frac{3}{10} + 0.375 \times \frac{4}{10}$

$= 0.416$

For weather - Gini Index = 0.416
Parents - " " = 0.36
Money - " " = 0.486

Out of these 3,
minimum Gini index is for parents
= 0.36 ⇒ max info gain
Select that root node.

Page No.

Date ___/___/___

Parents.

yes

no

possibilities - yes/no

Weekend	Weather	Parents	Money	Decision
W1	Sunny	Y	Rich	C
W3	Windy	Y	Rich	C
W4	Rainy	Y	Poor	C
W6	Rainy	Y	Poor	C
W9	Windy	Y	Rich	C

Week end	Weather	P	M	D
W2	Sunny	N	R	T
W5	Rainy	N	R	St
W7	Windy	N	P	C
W8	Windy	N	R	Sh
W10	Sunny	N	R	Te

For Parents = Yes
we get 1 set of data and the decision = Cinema

Parents - No
then Decision is a combination of Tennis, Cinema, Shopping, Stay in.

so will select one another attribute

Computation of Gini index for parents = No
Weather Attribute

Sunny - 2 examples with tennis.
For Parent = No
Gini (Sunny) = $1-\left(\frac{2}{2}\right)^2 = 0$

Rainy 1 example with stay in
Gini (Rainy) = $1-\left(\frac{1}{1}\right)^2 = 0$

Windy 1 time Cinema 1 time Shopping
Gini(Windy) = $1-\left[\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^2\right] = 0.5$

WAC(Parents = 0 | Weather) = $0*\frac{2}{5}+0*\frac{1}{5}+0.5*\frac{2}{5}$
$= 0.2$

Date

Money Attribute

Values: Rich (4 examples — 2 Tennis, 1 Shopping, 1 Stay in)

Poor (1 example - 1 cinema)

Computation of Gini Index for Parents = No | Money Attribute

- Poor (1 example)

For parents = No | Money = Poor, there is 1 example with Cinema

Gini (S) = $1 - \left(\frac{1}{1}\right)^2 = \underline{0}$

Rich (4 examples)

Gini (Rich) = $1 - \left(\left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right)$

$= \underline{0.625}$

Weighted Avg = $0.625 \times 4/5 + 0 \times 1/5$

$= 0.5$

Gini Index (Parents) ∞

Weather - 0.2

Money - 0.5

Weather is selected as next branch.

Date ___/___/_____

Now for Parent = No & Weather = Sunny we have all instances as Tennis

Weekend	Weather	Parents	Money	Decision
W2	Sunny	No	Rich	Tennis
W10	Sunny	No	Rich	Tennis.

so no need to split

Now for Parent = No & Weather = Rainy we have all instances as Stay In

Weekend	Weather	Parents	Money	Decision
W5	Rainy	No	Rich	StayIn

For Parent = No Weather = Windy

Weather	Weather	Parents	Money	Decision
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping

Parents = No Weather = Windy Money = Poor
⇒ Cinema