Gini Index

Dr. Parimala M, SCORE

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

Four possible output
Cinema- 6 instances
Tennis-2 instances
Stay In- 1 instances
Shopping- 1 instances

Formula of Gini Index

The formula of the Gini Index is as follows:

$$Gini = 1 - \sum_{i=1}^{n} (p_i)^2$$

where,

'pi' is the probability of an object being classified to a particular class.

While building the decision tree, we would prefer to choose the attribute/feature with the least Gini Index as the root node.

Computation of Gini Index for Money Attribute

It has two possible values of Rich (7 examples) and Poor (3 examples).

For **Money = Poor**, there are 3 examples with "Cinema".

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

For **Money = Rich**, there are 2 examples with "Tennis", 3 examples with "Cinema" and 1 example with "Stay in", "Shopping" each

$$Gini(S) = 1 - \left[\left(\frac{2}{7}\right)^2 + \left(\frac{3}{7}\right)^2 + \left(\frac{1}{7}\right)^2 + \left(\frac{1}{7}\right)^2\right] = 0.694$$

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

$$Weighted\ Average(Money)$$

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

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

Computation of **Gini Index for Parents** Attribute

It has two possible values of Yes (5 examples) and No (5 examples).

For **Parents = Yes**, there are 5 examples, all with "Cinema".

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

For **Parents = No**, there are 2 examples with "Tennis", 1 example with "Stay in", "Shopping" and "Cinema" each

$$Gini(S) = 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2\right] = 0.72$$

$$\textbf{Weighted Average(Parents)}$$

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

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

$$Weighted\ Average(Weather)$$

$$= 0.444 * \left(\frac{3}{10}\right) + 0.444 * \left(\frac{3}{10}\right) + 0.375 * \left(\frac{4}{10}\right)$$

$$= 0.416$$

Computation of **Gini Index for Weather** Attribute

It has three possible values of **Sunny (3 examples)**, **Rainy (3 examples)** and **Windy (4 examples)**.

For **Weather = Sunny**, there are **2 examples with "Cinema"** and **1 with "Tennis"**.

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

For **Weather = Rainy**, there are **2 examples with "Cinema"** and **1 example with "Stay in"**

For **Weather = Rainy**, there are **2 examples with "Cinema"** and **1 example with "Stay in"**

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

For **Weather = Windy**, there are **3 examples with "Cinema"** and **1 example with "Shopping"**

Gini(windy)= 1-[(3/4)^2+(1/4)^2]=0.375

For Weather - Gini Index: 0.416

For Parents - Gini Index: 0.36

For Money - Gini Index: 0.486

Parents is selected as it has smallest Gini index.

Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W6	Rainy	Yes	Poor	Cinema
W9	Windy	Yes	Rich	Cinema

Weekend	Weather	Parents	Money	Decision
W2	Sunny	No	Rich	Tennis
W5	Rainy	No	Rich	Stay In
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W10	Sunny	No	Rich	Tennis

Tree 1.1

Tree 1.2

Tree 1.2

Gini for Weather attribute

Weekend	Weather	Parents	Money	Decision
W2	Sunny	No	Rich	Tennis
W5	Rainy	No	Rich	Stay In
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W10	Sunny	No	Rich	Tennis

- **Sunny (2 examples)**
- For Parent= No |Weather = Sunny, there are 2 example with "Tennis.
- $Gini(S) = 1 - [(\frac{2}{2})^2] = 0$

- **Rainy (1 example).**
- For Parents = No | Weather = Rainy, there is 1 example with "Stay In".
- $Gini(S) = 1 - [(\frac{1}{1})^2] = 0$

- **Windy (2 example)**
- For Parents = No | Weather = Windy, there is 1 example with "Cinema" and 1 with "Shopping".
- $Gini(S) = 1 - \left[(\frac{1}{2})^2 + (\frac{1}{2})^2\right] = 0.5$

$$Weighted\ Average(Parents = No\,|\,Weather) = 0 * \left(\frac{2}{5}\right) + 0 * \left(\frac{1}{5}\right) + 0.5 * \left(\frac{2}{5}\right) = 0.2$$

Gini for Money attribute

- Rich (4 examples)
- For Parents = No| Money = Rich, there is 1 example with "stay in" and "Shopping" each and 2 examples of "Tennis".
- $Gini(S) = 1 - [\left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{2}{4}\right)^2] = 0.625$

- Poor (1 example)
- For Parents = No| Money = Poor, there is 1 example with "Cinema".
- $Gini(S) = 1 - [\left(\frac{1}{1}\right)^2] = 0$
- $Weighted\ Average\ (Parents = No|\ Money) = 0.625 * (4/5) + 0 * (1/5) = 0.5$

Weekend	Weather	Parents	Money	Decision
W2	Sunny	No	Rich	Tennis
W5	Rainy	No	Rich	Stay In
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W10	Sunny	No	Rich	Tennis

Weekend	Weather	Parents	Money	Decision
W2	Sunny	No	Rich	Tennis
W5	Rainy	No	Rich	Stay In
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W10	Sunny	No	Rich	Tennis

For Parents = No | Weather - Gini Index: 0.2

For Parents = No | Money - Gini Index: 0.5

Weather is selected as it has smallest Gini index.

Weekend	Weather	Parents	Money	Decision
W2	Sunny	No	Rich	Tennis
W5	Rainy	No	Rich	Stay In
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W10	Sunny	No	Rich	Tennis

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

Tree 2.1

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

Now, for Parents=No & Weather=Rainy, we have all instances as Stay In.

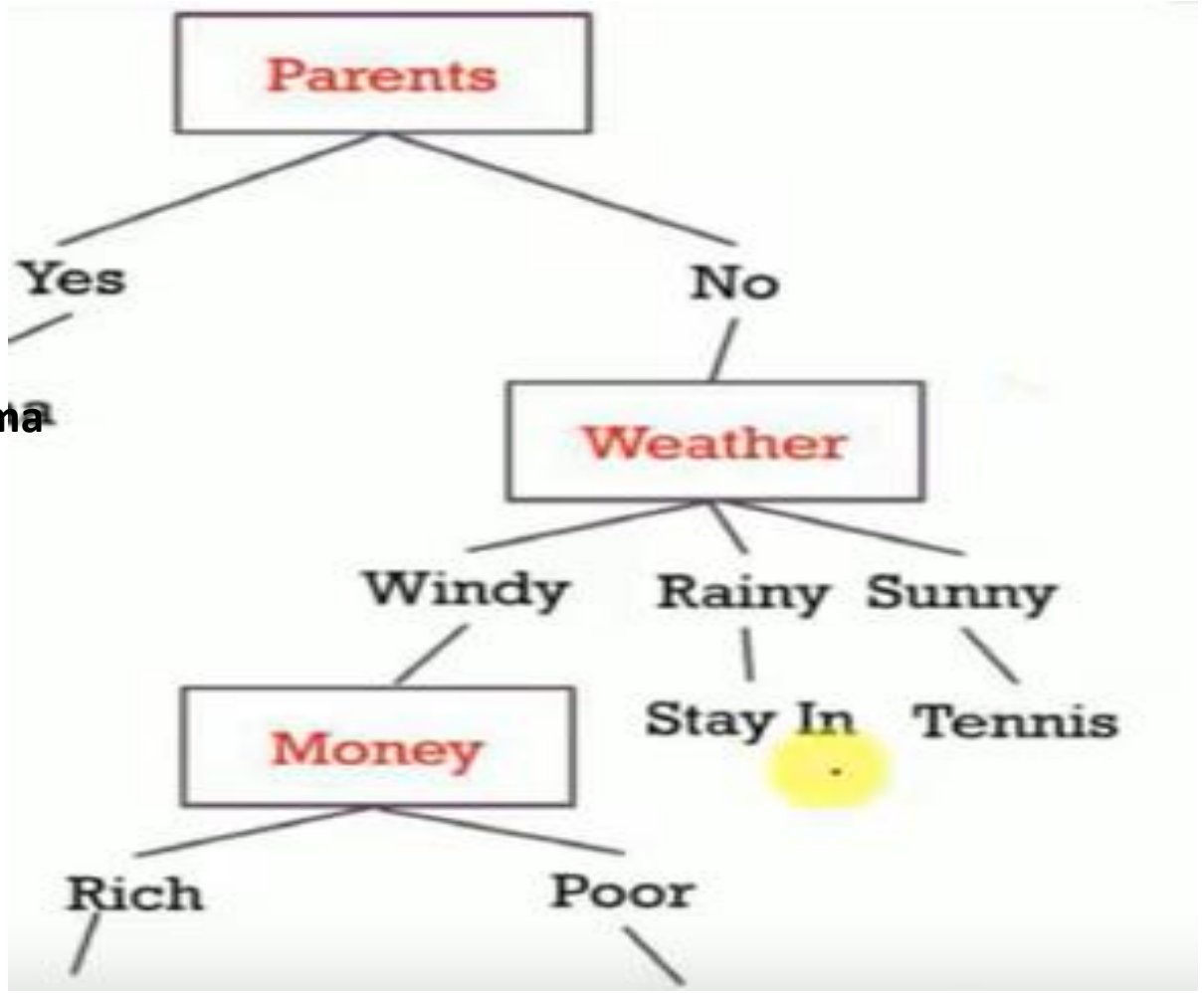
Tree 2.2

Weekend	Weather	Parents	Money	Decision
W5	Rainy	No	Rich	Stay In

Now, for Parent=No & Weather=Windy, we need to split.

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

Tree 2.3



cinema

shopping

cinema