

DATE: / /

## Assignment # 02

21-Arial-557

Name : Arslan Afzaal

Gini Index of class attribute:-

$$\text{Gini (overall class)} = 1 - \sum_{j=1}^n (p_j)^2$$

$$\text{Gini(class)} = 1 - \left(\frac{10}{20}\right)^2 - \left(\frac{10}{20}\right)^2 \Rightarrow 1 - 0.25 - 0.25$$

$$\boxed{\text{Gini(class)} = 0.5}$$

Compute the Gini index for the gender attribute:-

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

$$\text{Gini(male)} = 0.48$$

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

$$\boxed{\text{Gini(female)} = 0.48}$$

$$\text{Gini index (D)} = \frac{|D_1| \text{ gini}(D_1) + |D_2| \text{ gini}(D_2)}{|D|}$$

$$\text{Gini index(gender)} = \frac{10}{20} (0.48) + \frac{10}{20} (0.48)$$

$$\boxed{\text{Gini index(gender)} = 0.48}$$

→ Gini index for the car type attribute :

Gini(Family, sport, luxury)

The combination of these attributes are

 $\{( \text{family}, \text{sport}), (\text{family}, \text{luxury}), (\text{sport}, \text{luxury}),$   
 $(\text{family}), (\text{sport})(\text{luxury})\}$ 

gini(car type(Family,sport)) ∈ luxury

$$= \frac{12}{20} \left(1 - \left(\frac{9}{12}\right)^2 - \left(\frac{3}{12}\right)^2\right) + \frac{8}{20} \left(1 - \left(\frac{1}{8}\right)^2 - \left(\frac{7}{8}\right)^2\right)$$

DATE: 1/1

$$= \frac{12}{20}(0.375) + \frac{8}{20}(0.21875) \Rightarrow 0.3125$$

Gini (Family, Luxury &amp; Sport)

$$= \frac{12}{20} \left( 1 - \left( \frac{10}{12} \right)^2 - \left( \frac{2}{12} \right)^2 \right) + \frac{8}{20} \left( 1 - \left( \frac{8}{8} \right)^2 - \left( \frac{0}{8} \right)^2 \right)$$

$$= \frac{12}{20}(0.271) + \frac{8}{20}(0)$$

$$= 0.1662$$

Gini (Sport Luxury) &amp; Family :-

$$= \frac{16}{20} \left( 1 - \left( \frac{9}{16} \right)^2 - \left( \frac{7}{16} \right)^2 \right) + \frac{4}{20} \left( 1 - \left( \frac{1}{4} \right)^2 - \left( \frac{3}{4} \right)^2 \right)$$

$$= \frac{16}{20}(0.4921) + \frac{4}{20}(0.375)$$

$$= 0.4686$$

Gini index of shirt size attribute :-

Gini (small, medium, large, extra large).

→ Gini (small, medium) &amp; (large, extra large).

$$= \frac{12}{20} \left( 1 - \left( \frac{6}{12} \right)^2 - \left( \frac{6}{12} \right)^2 \right) + \frac{8}{20} \left( 1 - \left( \frac{4}{8} \right)^2 - \left( \frac{4}{8} \right)^2 \right)$$

$$= \frac{12}{20}(0.5) + \frac{8}{20}(0.5)$$

$$= 0.5$$

→ Gini (small, large) and (medium, extra large)

$$= \frac{9}{20} \left( 1 - \left( \frac{5}{9} \right)^2 - \left( \frac{4}{9} \right)^2 \right) + \frac{11}{20} \left( 1 - \left( \frac{8}{11} \right)^2 - \left( \frac{6}{11} \right)^2 \right)$$

$$= \frac{9}{20}(0.4938) + \frac{11}{20}(0.4958)$$

$$= 0.4949$$

→ Gini (small, extra-large) and (medium, large) :-

$$= \frac{9}{20} \left( 1 - \left( \frac{5}{9} \right)^2 - \left( \frac{4}{9} \right)^2 \right) + \frac{11}{20} \left( 1 - \left( \frac{3}{11} \right)^2 - \left( \frac{6}{11} \right)^2 \right)$$

$$= \frac{9}{20} (0.4938) + \frac{11}{20} (0.4958)$$

$$= 0.4949$$

→ Gini (medium, extra-large) or (small, extra-large) :-

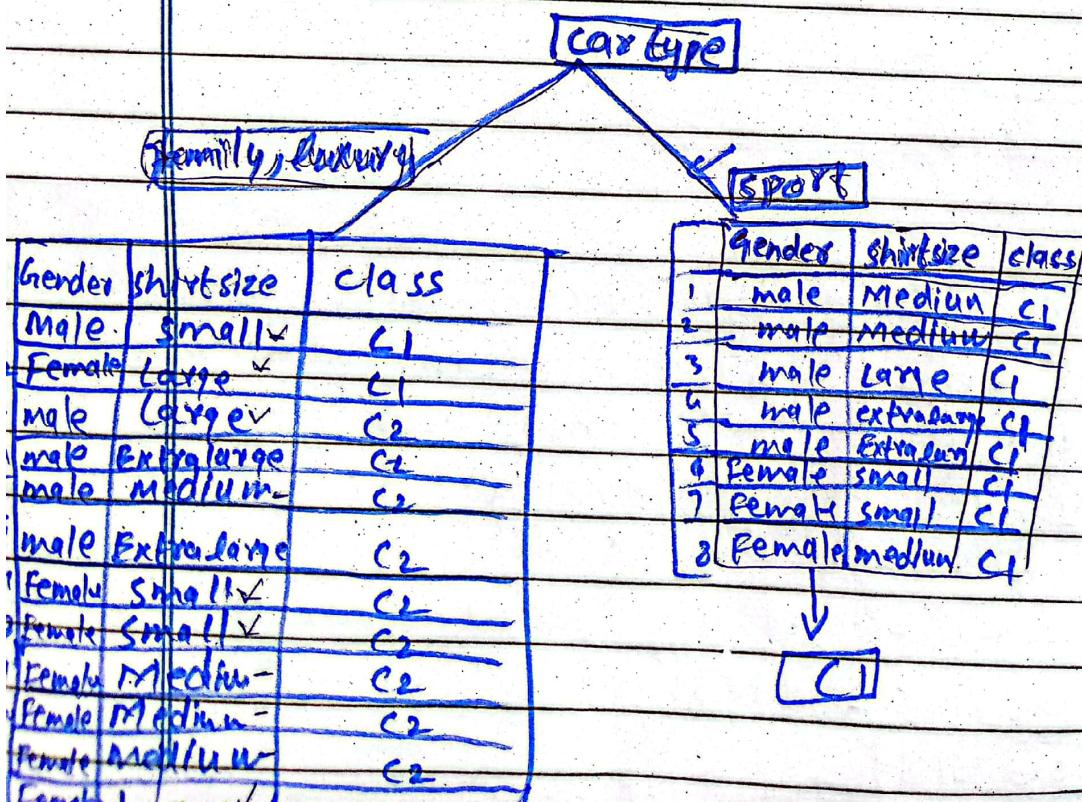
$$= \frac{11}{20} \left[ 1 - \left( \frac{5}{11} \right)^2 - \left( \frac{6}{11} \right)^2 \right] + \frac{9}{20} \left[ 1 - \left( \frac{3}{9} \right)^2 - \left( \frac{4}{9} \right)^2 \right]$$

$$= \frac{11}{20} (0.4958) + \frac{9}{20} (0.4938)$$

$$= 0.4949$$

The <sup>(Ans)</sup> Gini index of car types is less  
so car types is our root node.

Gini (family, luxury) and sport .



Gini of overall class:-

$$\text{Gini}_{\text{class}}(2,10) = 1 - \left(\frac{2}{12}\right)^2 - \left(\frac{10}{12}\right)^2$$

$$= 1 - 0.0277 - 0.6944$$

$$\boxed{\text{Gini(class)} = 0.6778}$$

Gini index of Gender:-

$$\text{Gini of male}(7,5) = 1 - \left(\frac{7}{12}\right)^2 - \left(\frac{5}{12}\right)^2 \Rightarrow 1 - 0.04 - 0.64$$

$$\boxed{\text{Gini(male)} = 0.32}$$

$$\text{Gini of female}(1,6) = 1 - \left(\frac{1}{7}\right)^2 - \left(\frac{6}{7}\right)^2 \Rightarrow 1 - 0.020408 - 0.73$$

$$\boxed{\text{Gini(female)} = 0.24483}$$

Gini index of Gender:-

$$\text{C.} = \frac{|D_1|}{|D|} \text{Gini(Male)} + \frac{|D_2|}{|D|} \text{Gini(Female)}$$

$$= \frac{5}{12} (0.32) + \frac{7}{12} (0.24483)$$

$$= 0.133 + 0.1428$$

$$\boxed{\text{Gini index(Gender)} = 0.2758}$$

Gini of Shirt size attribute:-

Gini(small, medium, large, extra-large)  
Gini(small, medium) and (large, extra-large):-

$$= \frac{7}{12} \left( \frac{1}{7} - \left(\frac{1}{7}\right)^2 - \left(\frac{6}{7}\right)^2 \right) + \frac{8}{12} \left( \frac{1}{8} - \left(\frac{1}{8}\right)^2 - \left(\frac{7}{8}\right)^2 \right)$$

$$= \frac{7}{12} (1 - 0.020408 - 0.73489) + \frac{5}{12} (1 - 0.04 - 0.64)$$

$$= 0.1428175 + 0.1333$$

$$\boxed{\leq 0.27615}$$

DATE: / /

Gini (small, large) and (medium, extra large)

$$= \frac{6}{12} \left[ 1 - \left( \frac{2}{8} \right)^2 - \left( \frac{4}{8} \right)^2 \right] + \frac{6}{12} \left[ 1 - \left( \frac{0}{6} \right)^2 - \left( \frac{6}{6} \right)^2 \right]$$

$$= \frac{6}{12} [1 - 0.25 - 0.44] + \frac{6}{12} (1 - 1)$$

$$= 0.5 (0.445)$$

Gini (small, large) = 0.2225

0  
medium  
extra large

Gini (small, extra large) and (medium, large)

$$= \frac{5}{12} \left[ 1 - \left( \frac{1}{5} \right)^2 - \left( \frac{4}{5} \right)^2 \right] + \frac{7}{12} \left[ 1 - \left( \frac{1}{7} \right)^2 - \left( \frac{6}{7} \right)^2 \right]$$

$$= \frac{5}{12} (1 - 0.04 - 0.64) + \frac{7}{12} (1 - 0.02048 - 0.73489)$$

$$= 0.27615$$

Gini (medium, extra large) and (small, large)

$$= \frac{6}{12} \left( \frac{10}{6} \right)^2 + \frac{6}{12} \left( \frac{12}{6} \right)^2 - \left( \frac{4}{8} \right)^2$$

$$= \frac{6}{12} (1 - 1) + \frac{6}{12} (1 - 0.25 - 0.44)$$

$$= 0.2225$$

shirt size is our root node

The Gini at p(small, large) and (medium, extra large) is less so they are our root node.

Cartype

Sport

Gender

Gender	Shirtsize	Class
Male	Small	C1
Female	Large	C1
male	Large	C2
male	Extra large	C2
male	Medium	C2
male	Extra large	C2
Female	Small	C2
Female	Small	C2
Female	Medium	C2
Female	Medium	C2
Female	Medium	C2
Female	Large	C2

Gender	Shirtsize	Class
Male	Medium	C1
male	Medium	C1
male	Large	C1
male	Extra large	C1
male	Extra large	C1
Female	Small	C1
Female	Small	C1
Female	Medium	C1

C1

Shirtsize

Small shirt

medium extra large

Gender Class
male C1
Female C1
male C2
Female C2
Female C2
Female C2

Gender Class
male C2
male C2
male C2
Female C2
Female C2

Female

C2



DATE: / /

### ~~Gini(class)~~ Gini(class), shirt size:-

$$\text{Gini(class)} = 1 - \left(\frac{2}{6}\right)^2 - \left(\frac{4}{6}\right)^2 \\ = 1 - 0.1111 - 0.44$$

$$\text{Gini(class)} = 1 - 0.53$$

$$\boxed{\text{Gini(class)} = 0.4445}$$

$$\text{Gini of male} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 \\ = 1 - 0.25 - 0.25$$

$$\boxed{\text{Gini male} = 0.5}$$

$$\text{Gini of female} = 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2 \\ = 1 - 0.0625 - 0.5625$$

$$\boxed{\text{Gini(female)} = 0.375}$$

Gini index of shirt sizes small, large

$$= \frac{2}{6}(0.5) + \frac{4}{6}(0.375)$$

$$= 0.166 + 0.25$$

$$\boxed{= 0.41}$$

Q No 3:- use the tree to classify  
the record female, family, small:-

ANSWER  
= Female  $\rightarrow$  Family  $\rightarrow$  small  $\rightarrow$  C<sub>1</sub>  
belongs to class(E1).