



HOMEWORK 4

DECISION TREE CLASSIFIER

Presented by: NMN & អុនុំយោបាយនារុន



MEMBER



นางสาวเพ็ญพิชชา วรรණ์ชุมາตร์
643020063-4



นายภักรธ ก้อนมณี
643020513-9



นายรัชชานนท์ กิพย์พิมานพร
643020515-5



นายอุษณิชัย คำนา
643020521-0



นางสาวริศรา ปันลา
643020519-7



นางสาวอนันต์ญา พูลสวัสดิ์
643020526-0



นางสาววิลันดา ทาระมาตย์
643021271-2

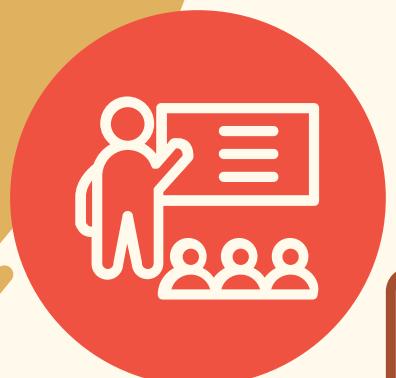
Parameters



Min_Samples_Split

เป็นพารามิเตอร์ที่ควบคุมจำนวนตัวอย่างขั้นต่ำที่ต้องอยู่ในโหนดได้ให้นدنี้ก่อนที่จะแบ่งโหนดนั้นต่อไป (โหนดจะไม่ถูกแบ่งหากมีจำนวนตัวอย่างน้อยกว่า)

จะแบ่งเป็น Decision node ก็ต่อเมื่อจำนวนของกลุ่มตัวอย่างมากกว่า หรือเท่ากับค่าที่กำหนดใน min samples split แต่ถ้าจำนวนของกลุ่มตัวอย่างน้อยกว่าค่าที่กำหนดจะไม่แบ่ง Decision node



Splitter

Best

หา feature ที่ดีที่สุด จากค่า entropy หรือ gini ดีที่สุด (ถ้าเป็นentropy ต้องการค่า gain สูงมากสุด) โดยมีแบ่ง threshold ไว้ค่าหนึ่งที่ไว้ใช้แบ่งซึ่งจะเป็นค่ากลางเสมอ เช่น 0.5 และทำการแบ่งข้อมูล

Random

จะมีการสุ่ม threshold ในทุกๆ feature ก่อน และค่อยหาค่า entropy หรือ gini ที่ดีที่สุด (ถ้าเป็นentropy ต้องการค่า gain สูงมากสุด) และทำการแบ่งข้อมูล

Splitter “Best”

Define

x0 = a9

x1 = a10

x2 = a12

x3 = a13

y = a16

Splitter “Best”

Define

$x_0 = a_9$

$x_1 = a_{10}$

$x_2 = a_{12}$

$x_3 = a_{13}$

$y = a_{16}$

Entropy

$$\text{Info}(D) = I(175, 149)$$

$$= \left[-\frac{175}{324} \log_2 \frac{175}{324} \right] + \left[-\frac{149}{324} \log_2 \frac{149}{324} \right]$$

$$= 0.479975 + 0.515375$$

$$= 0.99535 ,$$

$$\text{Gain}(a_9) = 0.99535 - 0.60845 = 0.3869$$

$$\text{Gain}(a_{10}) = 0.99535 - 0.84778 = 0.14757$$

$$\text{Gain}(a_{12}) = 0.99535 - 0.99442 = 0.00093$$

$$\text{Gain}(a_{13}) = 0.99535 - 0.98787 = 0.00748$$

$$\begin{array}{l} \text{Gain}(a_{13}) \xrightarrow{1,(2,3)} = 0.99535 - 0.993281 = 0.002069 \\ \quad \quad \quad \xrightarrow{(1,2),3} = 0.99535 - 0.989944 = 0.005406 \end{array}$$

Splitter “Best”

Define

$x_0 = a_9$

$x_1 = a_{10}$

$x_2 = a_{12}$

$x_3 = a_{13}$

$y = a_{16}$

Entropy

$$\text{Info}(D) = I(175, 149)$$

$$= \left[-\frac{175}{324} \log_2 \frac{175}{324} \right] + \left[-\frac{149}{324} \log_2 \frac{149}{324} \right]$$

$$= 0.479975 + 0.515375$$

$$= 0.99535,$$

$$\text{Gain}(a_9) = 0.99535 - 0.60845 = 0.3869$$

$$\text{Gain}(a_{10}) = 0.99535 - 0.84778 = 0.14757$$

$$\text{Gain}(a_{12}) = 0.99535 - 0.99442 = 0.00093$$

$$\text{Gain}(a_{13}) = 0.99535 - 0.98787 = 0.00748$$

$$\begin{aligned} \text{Gain}(a_{13}) &\rightarrow 1, (2, 3) = 0.99535 - 0.993281 = 0.002069 \\ &\rightarrow (1, 2), 3 = 0.99535 - 0.989944 = 0.005406 \end{aligned}$$

$$x(0) \leq 0.5$$

$$\text{entropy} = 0.995$$

$$\text{Sample} = 324$$

$$\text{Value} = [175, 149]$$

Splitter “Best”

กรณี $a_9 = 0$
หรือ $a_9 \leq 0.5$

$$\text{Info}(D) = I(133, 11)$$

$$\begin{aligned} &= \left[-\frac{133}{144} \log_2 \frac{133}{144} \right] + \left[-\frac{11}{144} \log_2 \frac{11}{144} \right] \\ &= 0.28344 + 0.105885 \\ &= 0.389325 \end{aligned}$$

$$\text{Gain}(a_{10}) = 0.389325 - 0.363416 = 0.025909$$

$$\text{Gain}(a_{12}) = 0.389325 - 0.383197 = 0.000128$$

$$\begin{aligned} \text{Gain}(a_{13}) &\rightarrow = 0.389325 - 0.31764 = 0.071685 \rightarrow \text{กรณี } a_{13} \text{ ก็แบ่งเป็นกรณี } 1, (2,3) \\ &\quad \swarrow \\ &= 0.389325 - 0.378532 = 0.010793 \end{aligned}$$

Splitter “Best”

$$a_9 = 0$$

$$\text{Info}(D) = I(133, 11)$$

$$\begin{aligned}
 &= \left[-\frac{133}{144} \log_2 \frac{133}{144} \right] + \left[-\frac{11}{144} \log_2 \frac{11}{144} \right] \\
 &= 0.28344 + 0.105885 \\
 &= 0.389325
 \end{aligned}$$

$$\text{Gain}(a_{10}) = 0.389325 - 0.363416 = 0.025909$$

$$\text{Gain}(a_{12}) = 0.389325 - 0.383197 = 0.000128$$

$$\begin{aligned}
 \text{Gain}(a_{13}) &\rightarrow = 0.389325 - 0.31764 = 0.071685 \rightarrow \text{กรณี } a_{13} \text{ ก็แบ่งเป็นกรณี } 1, (2,3) \\
 &= 0.389325 - 0.378532 = 0.010793
 \end{aligned}$$

$$x(0) \leq 0.5$$

$$\text{entropy} = 0.995$$

$$\text{Sample} = 324$$

$$\text{Value} = [175, 149]$$

$$x(3) \leq 1.5$$

$$\text{entropy} = 0.389325$$

$$\text{Sample} = 144$$

$$\text{Value} = [133, 11]$$

Splitter “Best”

กรณี $\theta_9 = 1$
หรือ $\theta_9 \geq 0.5$

$$\text{Info}(D) = I(42, 138)$$

$$\begin{aligned}
 &= \left[-\frac{42}{180} \log_2 \frac{42}{180} \right] + \left[-\frac{138}{180} \log_2 \frac{138}{180} \right] \\
 &= 0.489892 + 0.293885 \\
 &= 0.783777
 \end{aligned}$$

$\text{Gain}(\theta_{10}) = 0.783777 - 0.667931$	$= 0.115846$
$\text{Gain}(\theta_{12}) = 0.783777 - 0.759011$	$= 0.0024766$
$\text{Gain}(\theta_{13}) = 0.783777 - 0.7726283$	$= 0.0111487$

Splitter “Best”

กรณี $a_9 = 1$
หรือ $a_9 \geq 0.5$

$$\text{Info}(D) = I(42, 138)$$

$$= \left[-\frac{42}{180} \log_2 \frac{42}{180} \right] + \left[-\frac{138}{180} \log_2 \frac{138}{180} \right]$$

$$= 0.489892 + 0.293885$$

$$= 0.783777 \approx$$

$$x(0) \leq 0.5$$

$$\text{entropy} = 0.995$$

$$\text{Sample} = 324$$

$$\text{Value} = [175, 149]$$

$$x(3) \leq 1.5$$

$$\text{entropy} = 0.389325$$

$$\text{Sample} = 144$$

$$\text{Value} = [133, 11]$$

$$x(1) \leq 0.5$$

$$\text{entropy} = 0.783777$$

$$\text{Sample} = 180$$

$$\text{Value} = [42, 138]$$

$$\text{Gain}(a_{10}) = 0.783777 - 0.667931 = 0.115846$$

$$\text{Gain}(a_{12}) = 0.783777 - 0.759011 = 0.024766$$

$$\text{Gain}(a_{13}) = 0.783777 - 0.7726283 = 0.0111487$$

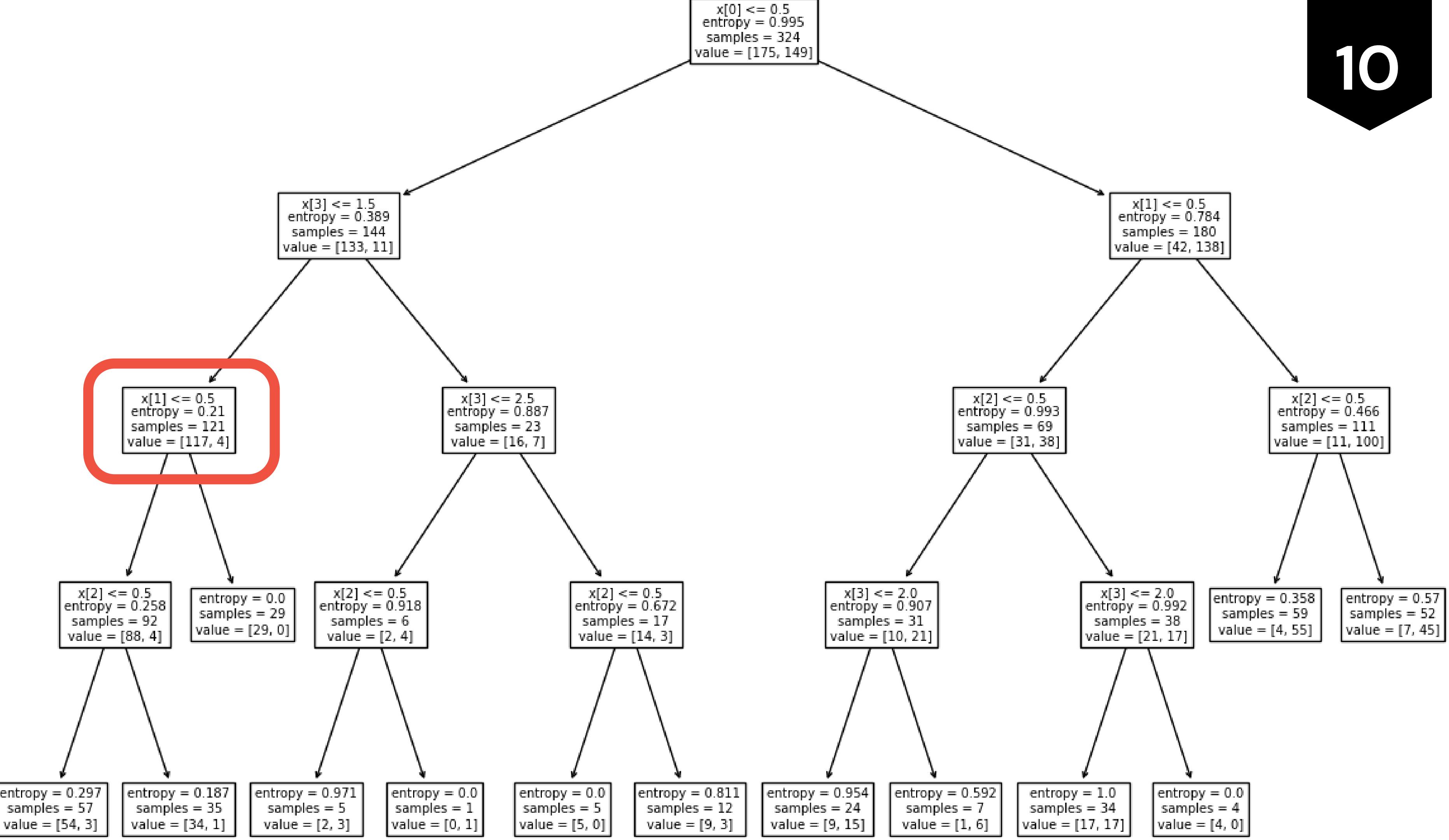
Splitter “Best”

กรณี $a_9 = 0$ ॥
หรือ $a_9 \leq 0.5$ ॥
และ $a_{13} = 1$
หรือ $a_{13} \leq 1.5$

$$\text{Info}(D) = I(117, 4)$$

$$\begin{aligned}
 &= \left[-\frac{117}{121} \log_2 \frac{117}{121} \right] + \left[-\frac{4}{121} \log_2 \frac{4}{121} \right] \\
 &= 0.0468953 + 0.162607 \\
 &= 0.2095023 //
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(a_{10}) &= 0.2095023 - 0.19618 &= 0.0133223 \\
 \text{Gain}(a_{12}) &= 0.2095023 - 0.2068932 &= 0.0026091
 \end{aligned}$$



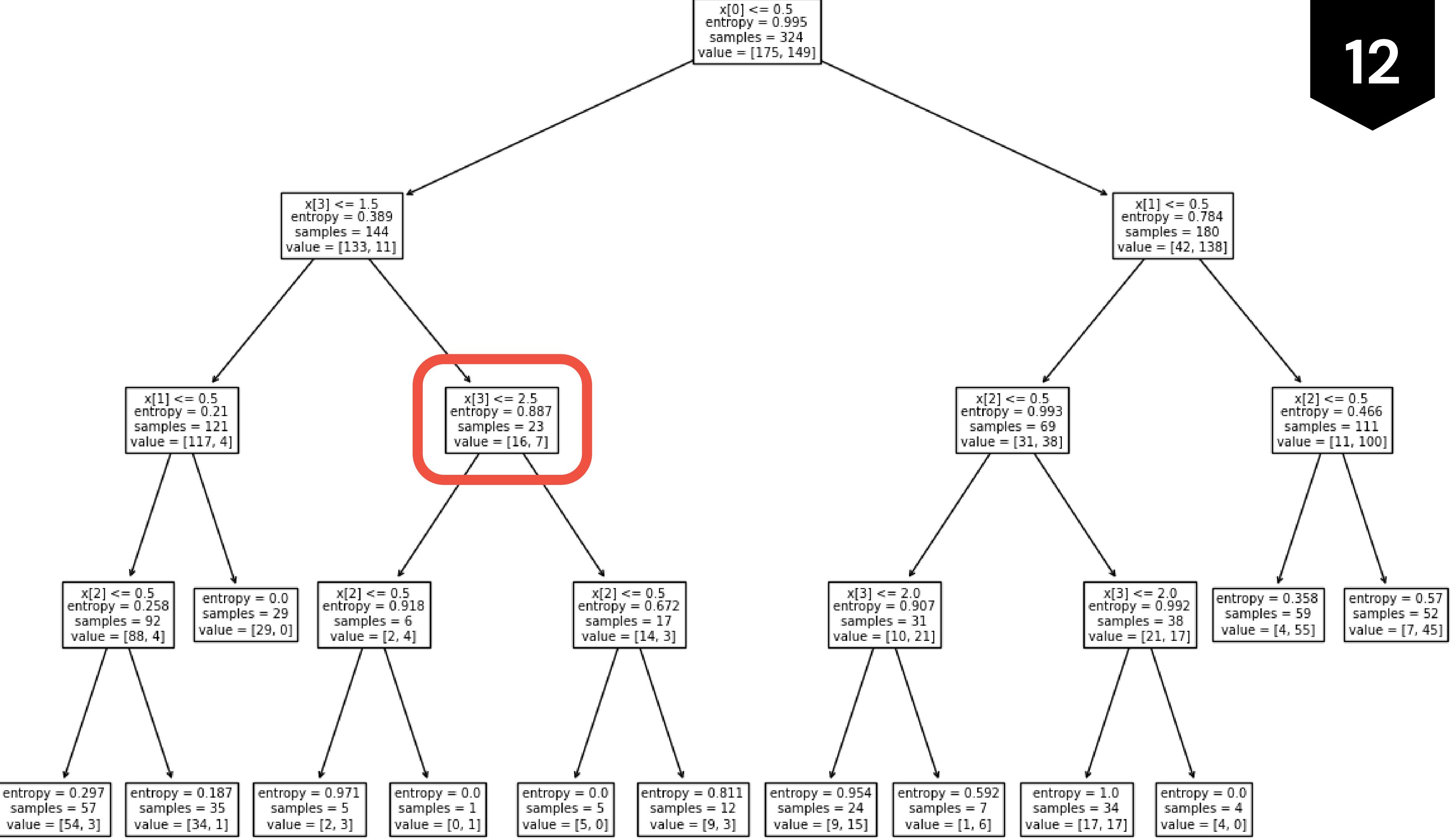
Best Samples Split

กรณี $a_9 = 0$ ให้ $a_{13} = 2, 3$
 หรือ $a_9 \leq 0.5$ ให้ $a_{13} \geq 1.5$

$$\text{Info}(D) = I(16, 7)$$

$$\begin{aligned}
 &= \left[-\frac{16}{23} \log_2 \frac{16}{23} \right] + \left[-\frac{7}{23} \log_2 \frac{7}{23} \right] \\
 &= 0.364217 + 0.522324 \\
 &= 0.886541 //
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(a_{10}) &= 0.886541 - 0.886541 = 0 \\
 \text{Gain}(a_{12}) &= 0.886541 - 0.88641 = 0.000051 \\
 \text{Gain}(a_{13}) &= 0.886541 - 0.736469 = 0.150072
 \end{aligned}$$



Splitter “Best”

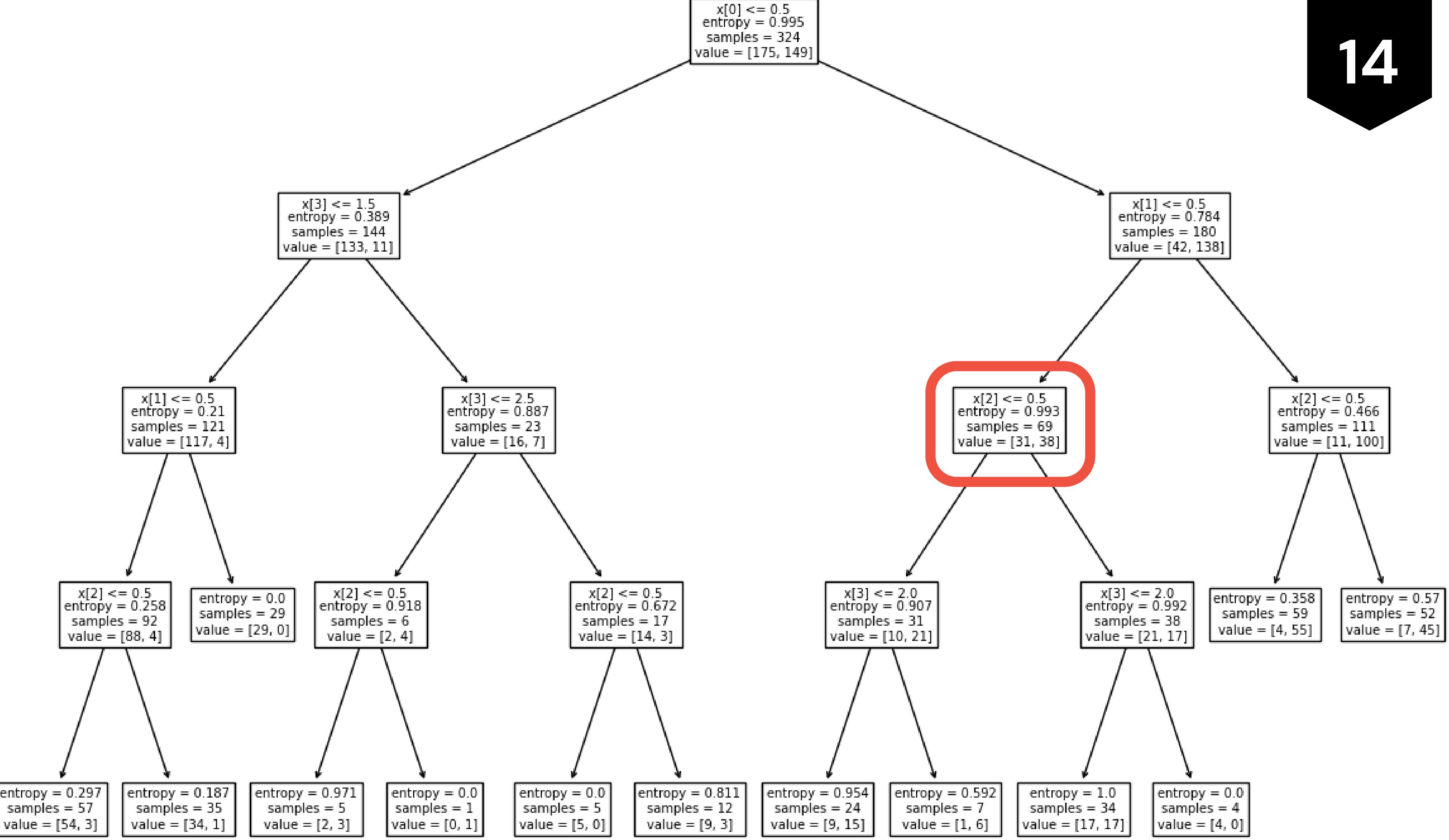
กรณี $a_9 = 1$ ให้ $a_{10} = 0$
 หรือ $a_9 \geq 0.5$ ให้ $a_{10} \leq 0.5$

$$\text{Info}(D) = I(31, 38)$$

$$= \left[-\frac{31}{69} \log_2 \frac{31}{69} \right] + \left[-\frac{38}{69} \log_2 \frac{38}{69} \right]$$

$$= 0.992563 \approx$$

$$\begin{aligned} \text{Gain}(a_{12}) &= 0.992563 - 0.953881 = 0.038682 \\ \text{Gain}(a_{13}) &= 0.992563 - 0.992549 = 0.000014 \end{aligned}$$



Splitter “Best”

กรณี $\hat{a}_9 = 1$ ॥ 亦 $\hat{a}_{10} = 1$
 หรือ $\hat{a}_9 \geq 0.5$ ॥ 亦 $\hat{a}_{10} \geq 0.5$

$$\text{Info}(D) = I(11, 100)$$

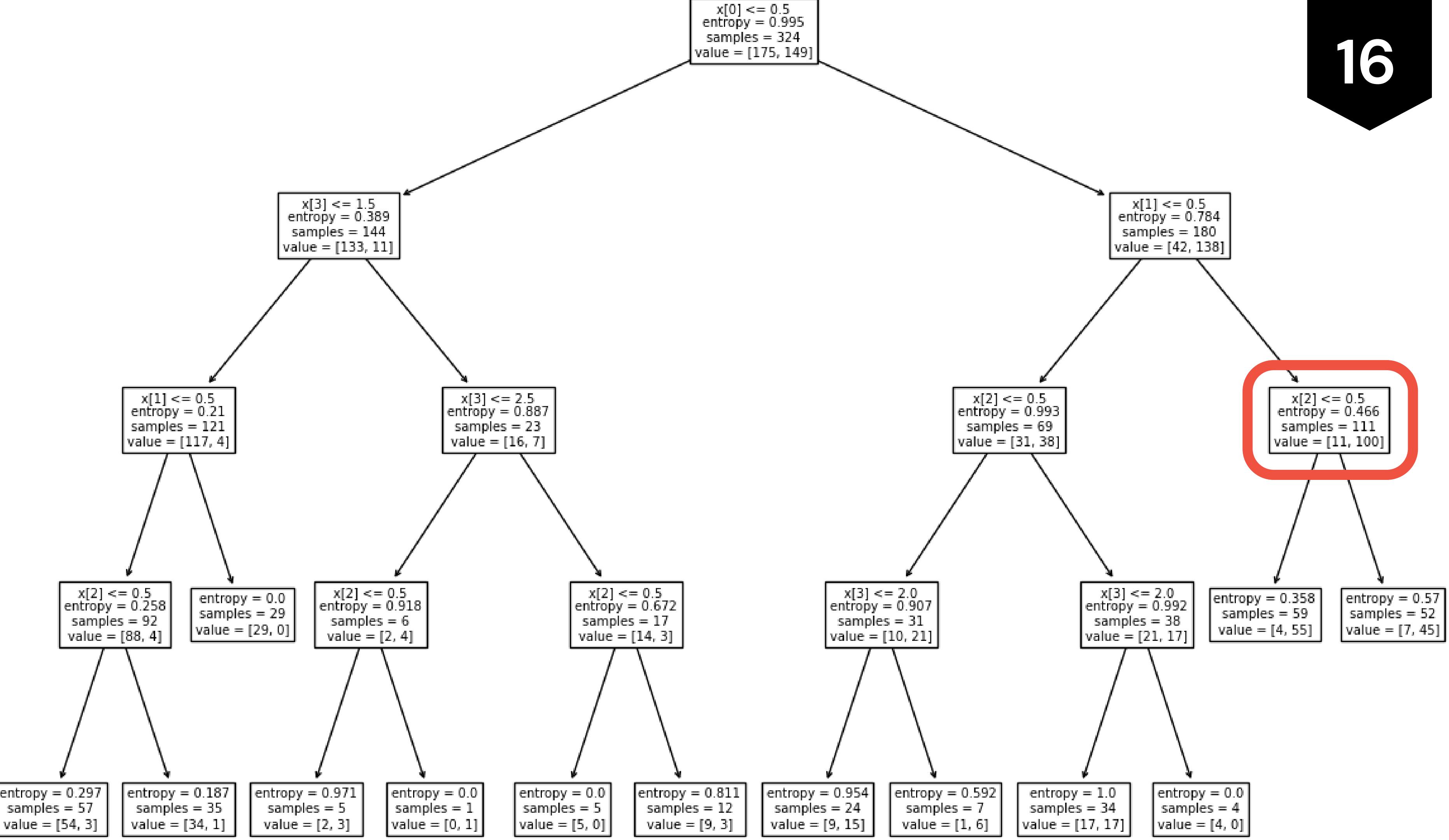
$$= \left[-\frac{11}{111} \log_2 \frac{11}{111} \right] + \left[-\frac{100}{111} \log_2 \frac{100}{111} \right]$$

$$= 0.330494 + 0.135639$$

$$= 0.466133,$$

$$\text{Gain}(\hat{a}_{12}) = 0.466133 - 0.457109 = 0.009024$$

$$\text{Gain}(\hat{a}_{13}) = 0.466133 - 0.466133 = 0$$



Splitter “Best”

గ්‍රන් $a_9 = 0$, $a_{13} = 1$, $a_{10} = 0$
 න්‍යෝ $a_9 \leq 0.5$, $a_{13} \leq 1.5$, $a_{10} \leq 0.5$

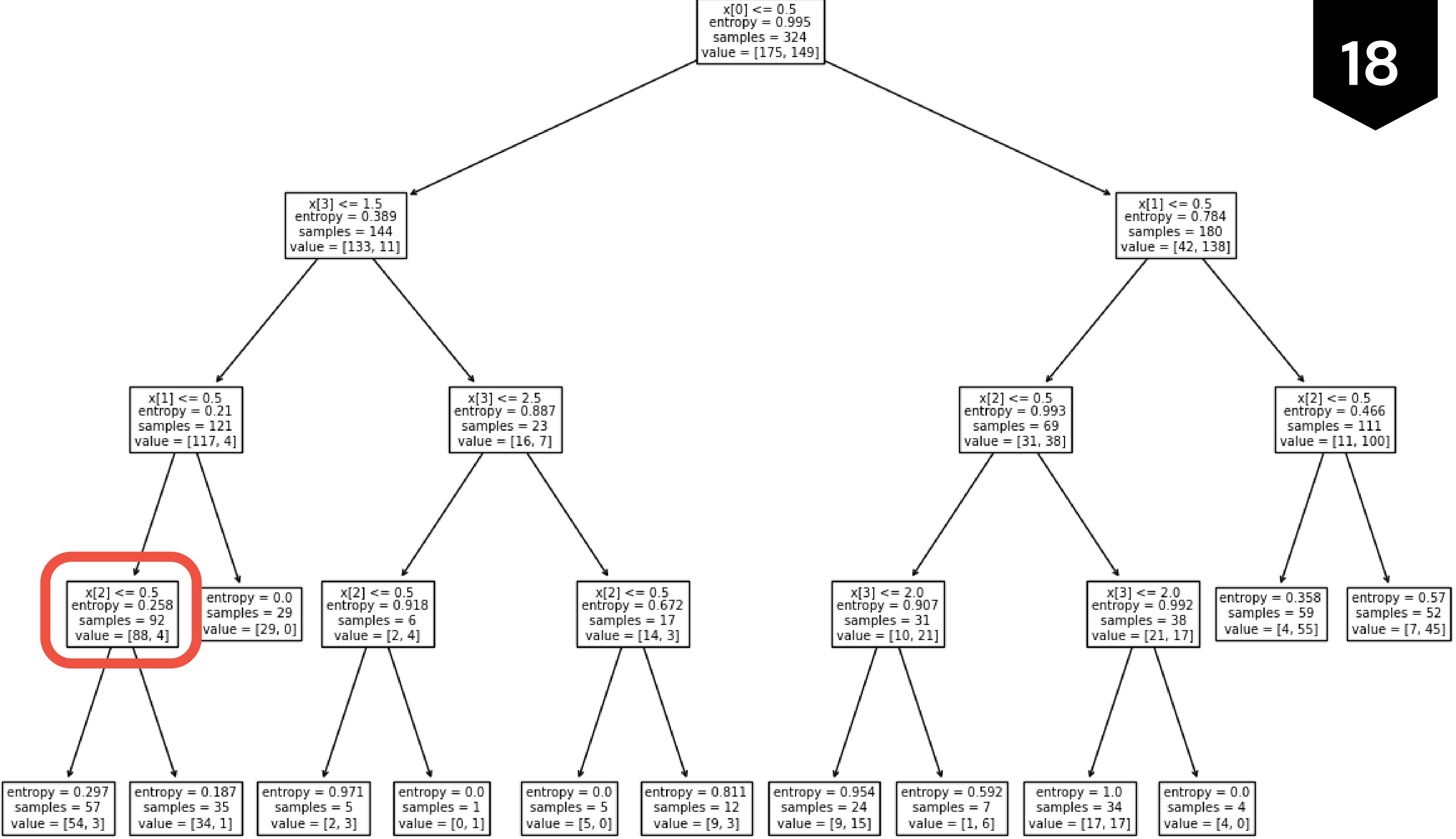
$$\text{Info}(D) = I(88, 4)$$

$$= \left[-\frac{88}{92} \log_2 \frac{88}{92} \right] + \left[-\frac{4}{92} \log_2 \frac{4}{92} \right] = 0.258019 //$$

$$\text{Gain}(a_{12}) = 0.258019 - 0.2555113 = 0.0025077$$

Column Labels

	0	1	Grand Total
	54	3	57
	34	1	35
	88	4	92

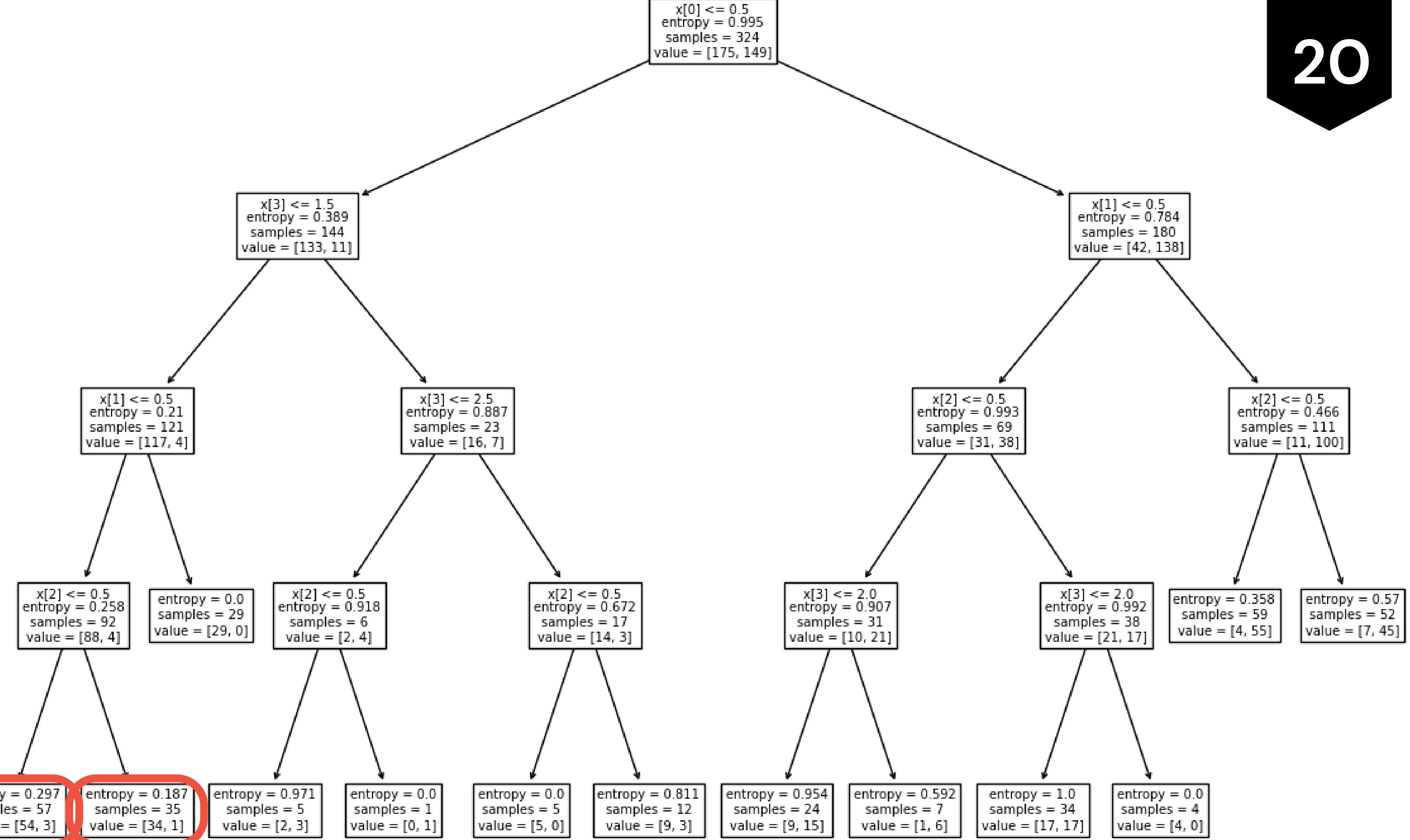


Splitter “Best”

$$\begin{aligned}
 &= 0.258019 // \\
 \text{Info}_{a_2}(D) &= \frac{57}{92} I(54,3) + \frac{35}{92} I(34,1) \\
 &= \frac{57}{92} \left[-\frac{54}{57} \log_2 \frac{54}{57} + \left(-\frac{3}{57} \log_2 \frac{3}{57} \right) \right] + \frac{35}{92} \left[-\frac{34}{35} \log_2 \frac{34}{35} + \left(-\frac{1}{35} \log_2 \frac{1}{35} \right) \right] \\
 &= 0.184303 + 0.0712083 \\
 &= 0.2555113 //
 \end{aligned}$$

▶ សំណុល់តាមការប្រើ feature, column, dimension, attribute គ្មាន។

Column Labels	0	1	Grand Total
	54	3	57
	34	1	35
	88	4	92



Splitter “Best”

กรณี $a_9 = 0$, $a_{13} = 1$, $a_{10} = 1$
 นั่นคือ $a_9 \leq 0.5$, $a_{13} \leq 1.5$, $a_{10} \geq 0.5$

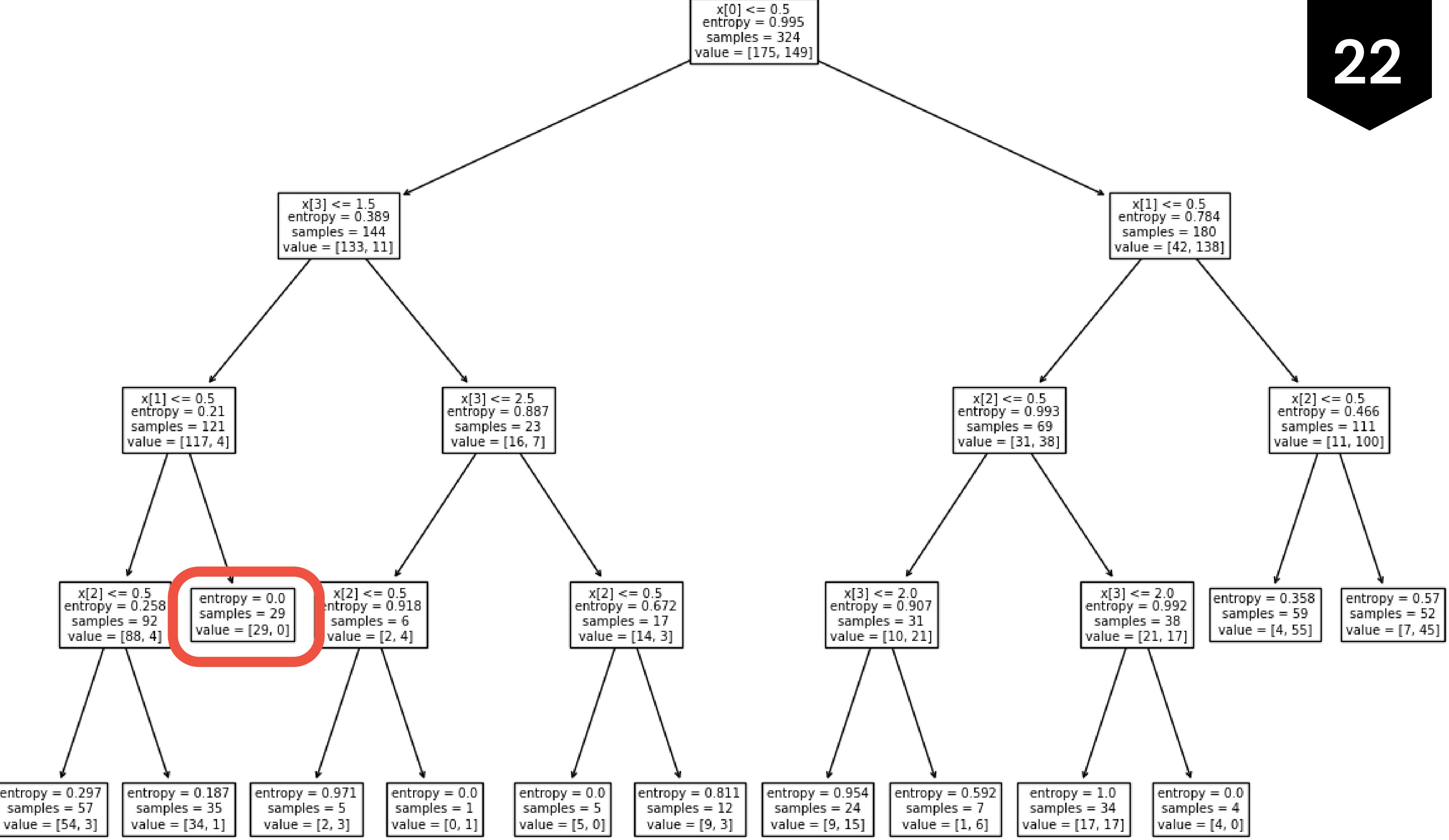
$$\text{Info}(D) = I(15, 14)$$

$$= \left[-\frac{15}{29} \log_2 \frac{15}{29} \right] + \left[-\frac{14}{29} \log_2 \frac{14}{29} \right]$$

$$= 0.999142 ,$$

Count of a Column Labels				
Row La	0	15	14	29
0				
Grand Tot	15	14	29	

$$\text{Gain}(a_{12}) = 0.999142 - 0.999142 = 0$$



Splitter “Best”

กรณี $\partial_9 = 0$ ให้ $\partial_{13} = (2,3)$ นิม $\partial_{13} = 2$
 หรือ $\partial_9 \leq 0.5$ ให้ $\partial_{13} \geq 1.5$ นิม $\partial_{13} \leq 2.5$

$$\text{Info}(D) = I(2,4)$$

$$= \left[-\frac{2}{6} \log_2 \frac{2}{6} \right] + \left[-\frac{4}{6} \log_2 \frac{4}{6} \right]$$

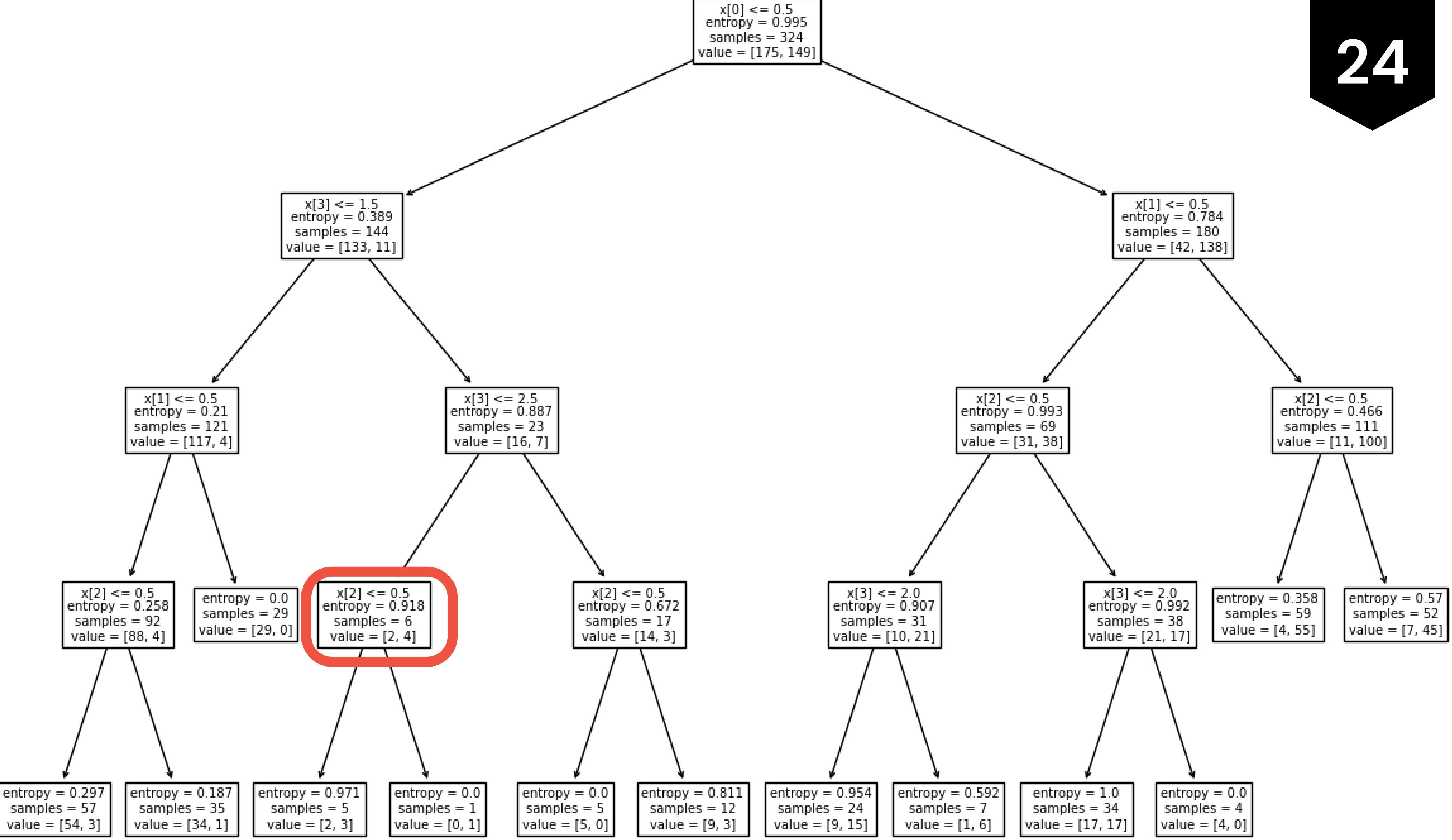
$$= 0.918296,$$

↗ 0

Column Labels

	0	1	Grand Total
2	3	1	5
1	1	1	2
2	4	4	6

$$\text{Gain}(\partial_{12}) = 0.918296 - 0.809126 = 0.10917$$



Splitter “Best”

$$= 0.918296 \text{ n}$$

$$\text{Info}_{a_{12}}(D) = \frac{5}{6} I(2,3) + \frac{1}{6} I(0,1)$$

$$= \frac{5}{6} \left[-\frac{2}{5} \log_2 \frac{2}{5} + \left(-\frac{3}{5} \log_2 \frac{3}{5} \right) \right]$$

$$= 0.809126 \text{ n}$$

0.97095

~~min 1/10~~

0

$$+ \frac{1}{6} \left[-\frac{0}{1} \log_2 \frac{0}{1} + \left(-\frac{1}{1} \log_2 \frac{1}{1} \right) \right]$$

Count of Column Labels

Row La



0 Grand Total

0

0

5

1

9

3

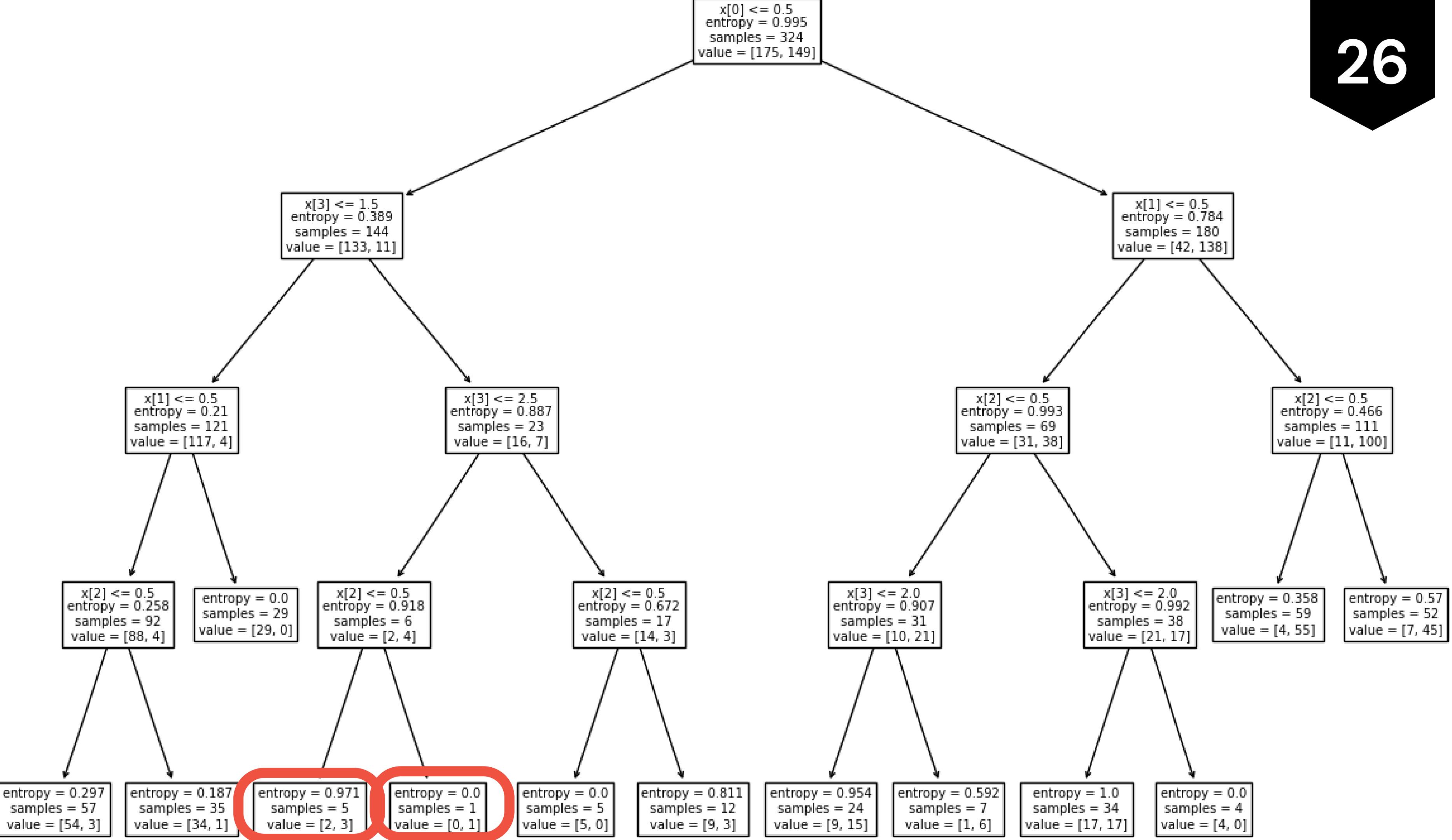
12

Grand Tot

14

3

17



Splitter “Best”

กรณี $a_9 = 0$, $a_{13} = (2,3)$ 乃是 $a_{13} = 3$
 หรือ $a_9 \leq 0.5$; $a_{13} \geq 1.5$ 乃是 $a_{13} \geq 2.5$

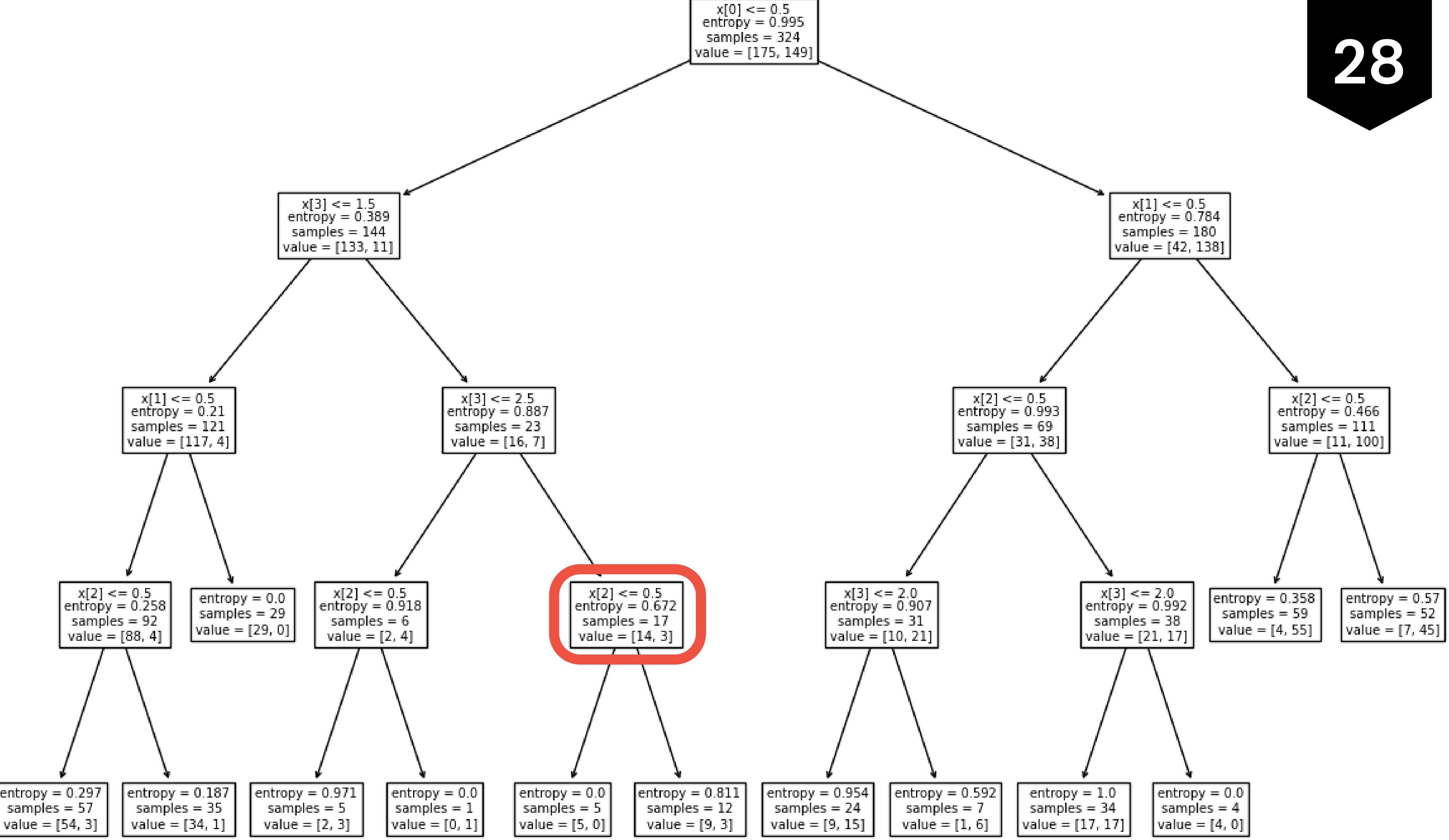
$$\text{Info}(D) = I(14, 3)$$

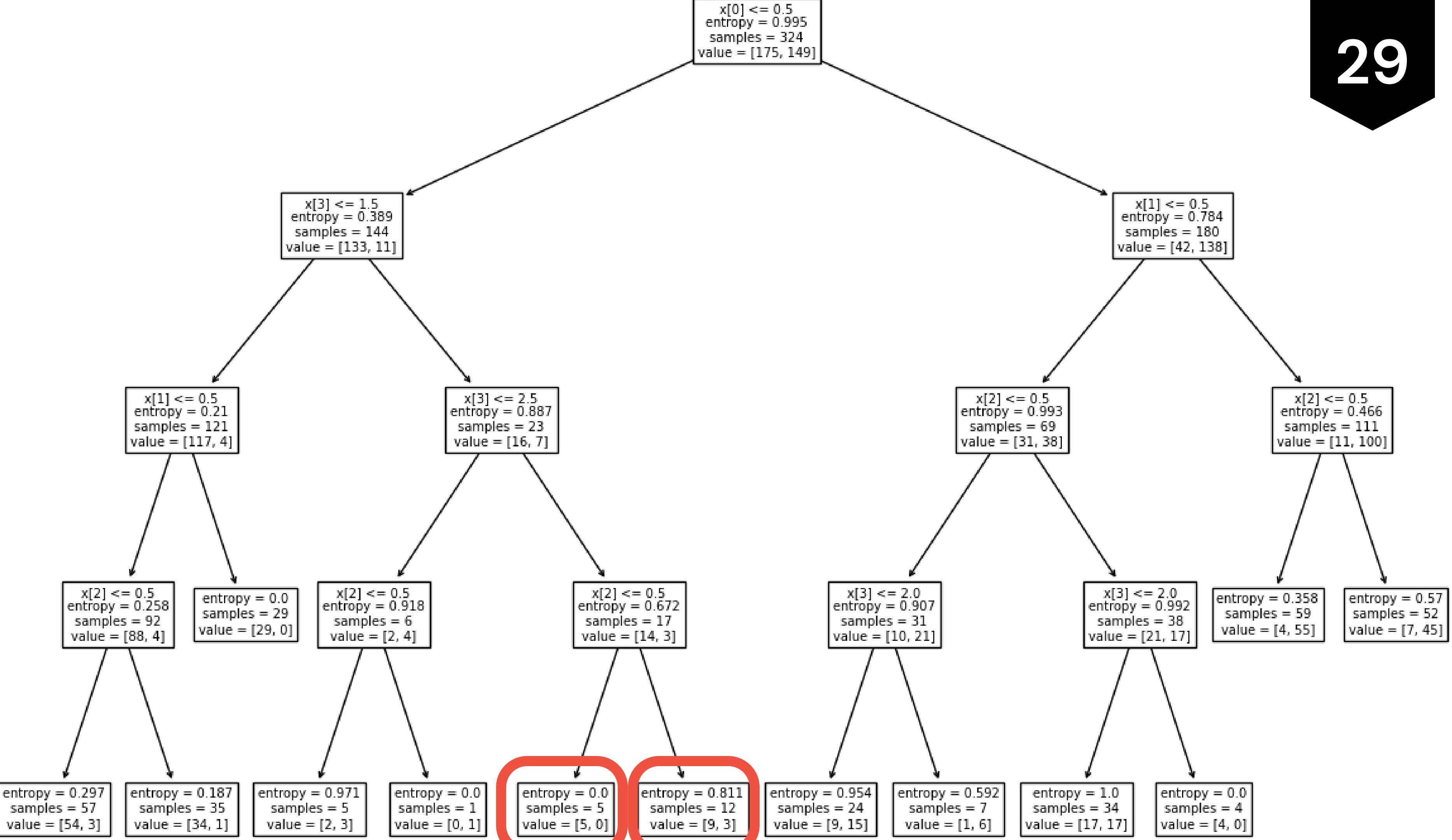
$$= \left[-\frac{14}{17} \log_2 \frac{14}{17} \right] + \left[-\frac{3}{17} \log_2 \frac{3}{17} \right]$$

$$= 0.672295 //$$

ดูแล้ว

$$\text{Gain}(a_{12}) = 0.672295 - 0.57667 = 0.095628$$





Splitter “Best”

)

$$\text{กรณี } \partial_9 = 1, \quad \partial_{10} = 0, \quad \partial_{12} = 0 \\ \text{ทั่วไป } \partial_9 \geq 0.5, \quad \partial_{10} \leq 0.5, \quad \partial_{12} \leq 0.5$$

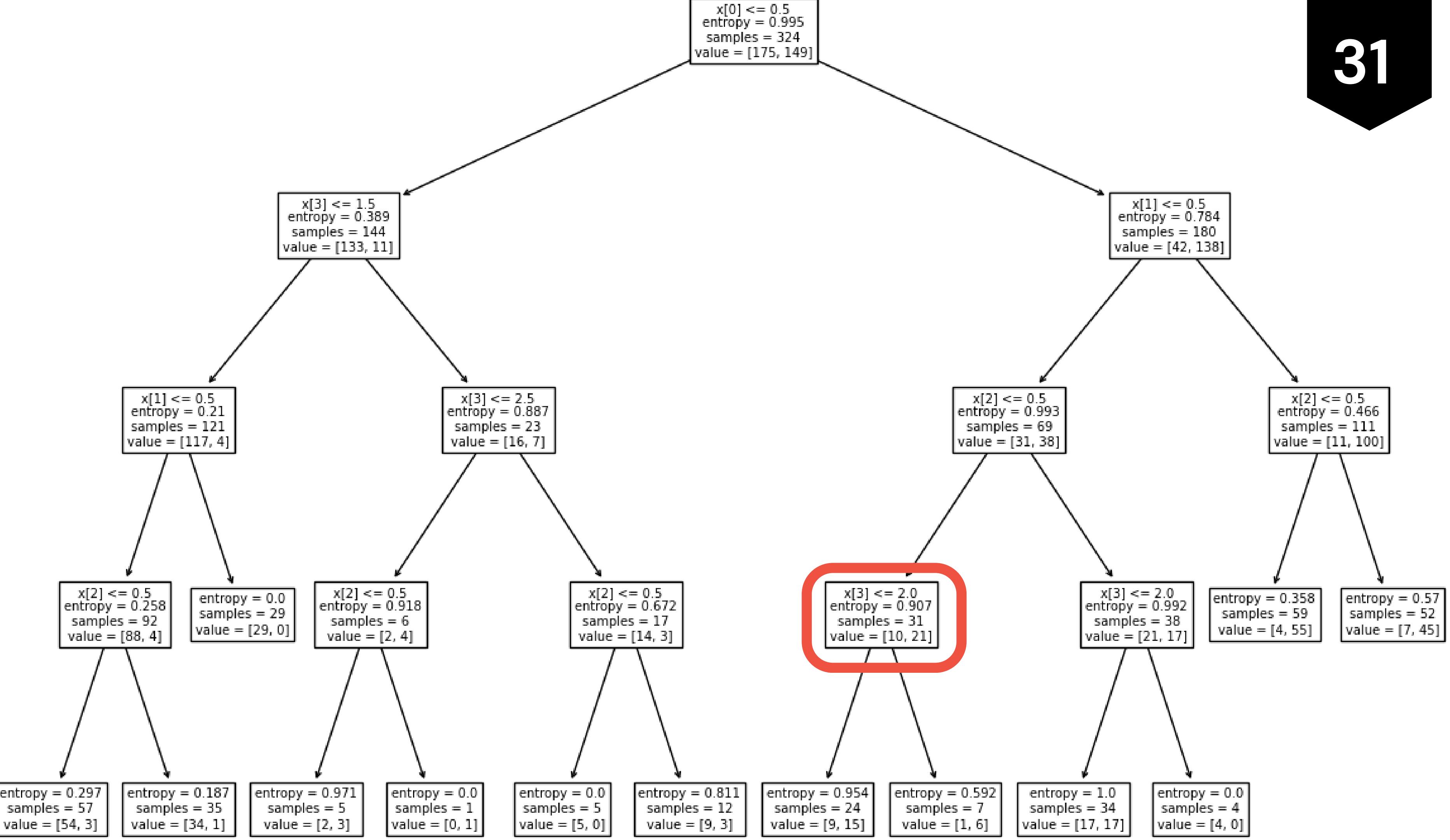
$$\text{Info}(D) = I(10, 21)$$

$$= \left[-\frac{10}{31} \log_2 \frac{10}{31} \right] + \left[-\frac{21}{31} \log_2 \frac{21}{31} \right] \\ = 0.526538 + 0.380628 \\ = 0.907166,$$

Column Labels

	0	1	Grand Total
4	55	59	59
4	55	59	59

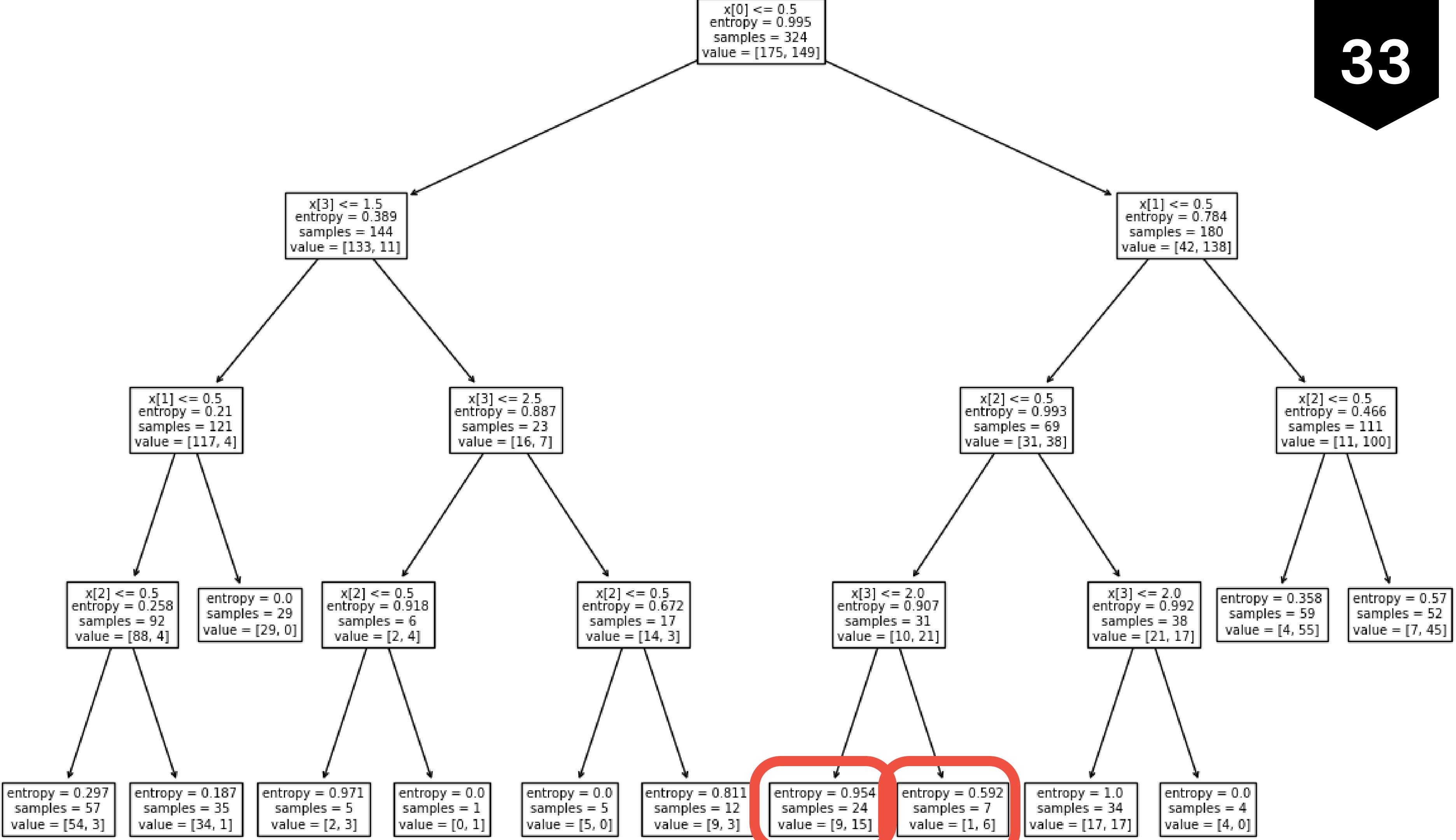
$$\text{Gain } (\partial_{13}) = 0.907166 - 0.872521 = 0.034645$$



Splitter “Best”

$$\begin{aligned}
 \text{Info } a_{13}(D) &= \frac{24}{31} \boxed{I(9, 15)} + \frac{7}{31} \boxed{I(1, 6)} \\
 &= \frac{24}{31} \left[-\frac{9}{24} \log_2 \frac{9}{24} + \left(-\frac{15}{24} \log_2 \frac{15}{24} \right) \right] + \frac{7}{31} \left[-\frac{1}{7} \log_2 \frac{1}{7} + \left(-\frac{6}{7} \log_2 \frac{6}{7} \right) \right] \\
 &= 0.738917 + 0.133604 \\
 &= 0.872521 //
 \end{aligned}$$

$$\text{Gain}(a_{13}) = 0.907166 - 0.872521 = \boxed{0.034645}$$



Splitter “Best”

กรณี $\theta_9 = 1, \theta_{10} = 0, \theta_{12} = 1$
นั่นคือ $\theta_9 \geq 0.5, \theta_{10} \leq 0.5, \theta_{12} \geq 0.5$

$$\text{Info}(D) = I(21, 17)$$

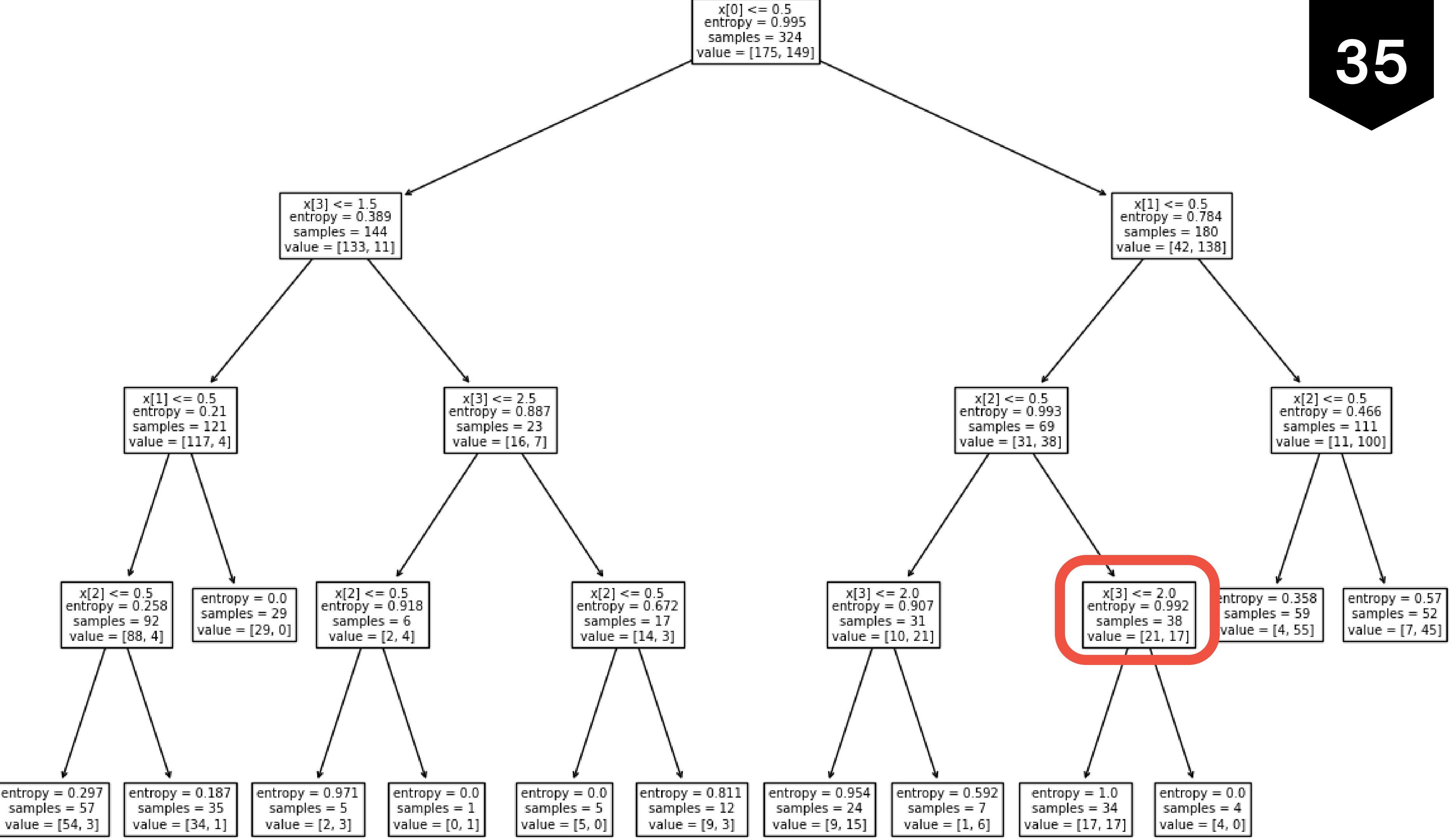
$$= \left[-\frac{21}{38} \log_2 \frac{21}{38} \right] + \left[-\frac{17}{38} \log_2 \frac{17}{38} \right]$$

$$= 0.472837 + 0.519155$$

$$= 0.991992 //$$

Row Labels	Column Labels		Grand Total
	0	1	
0	10	21	31
1	21	17	38
Grand Total	31	38	69

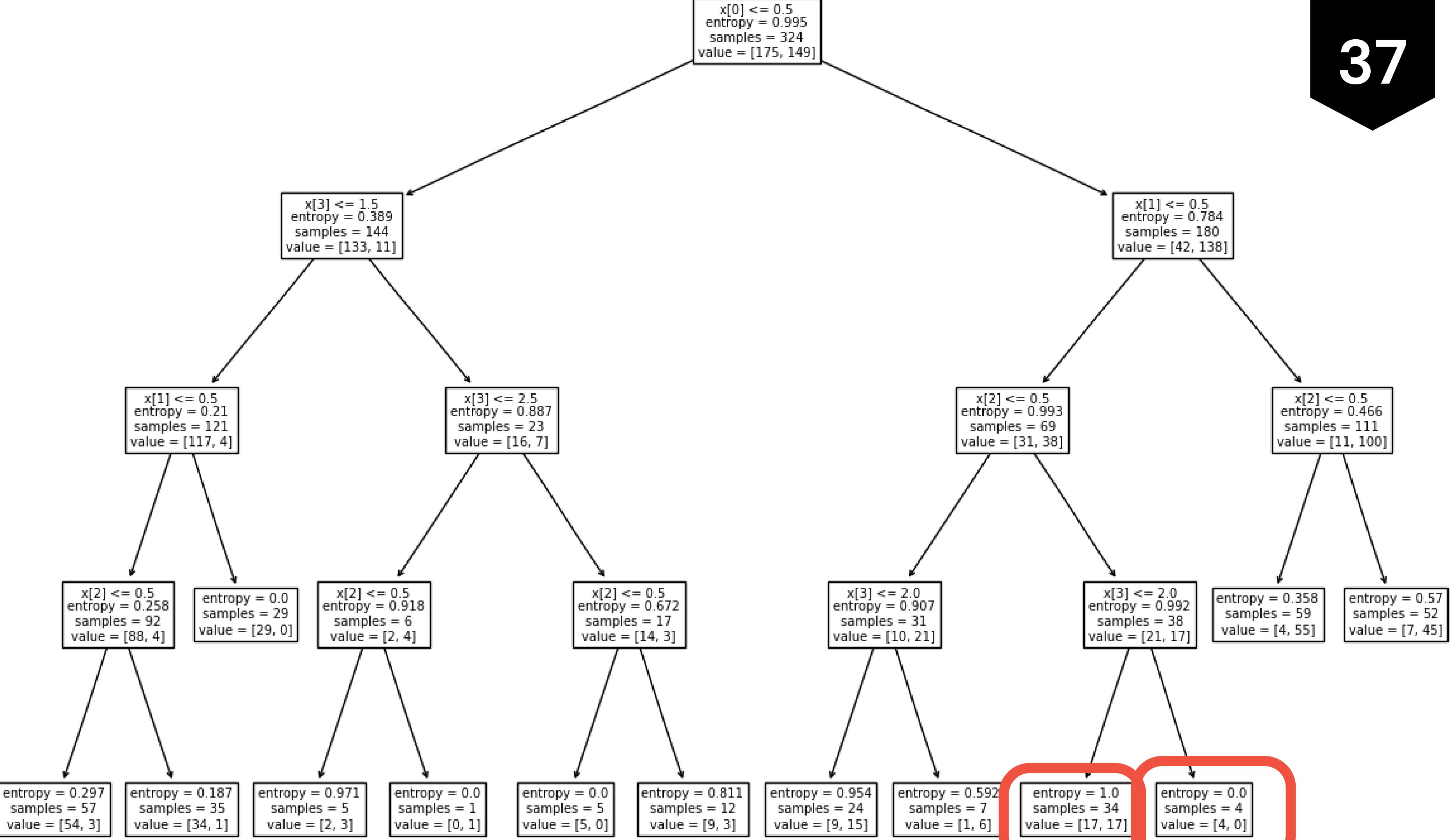
$$\text{Gain}(\theta_{13}) = 0.991992 - 0.894737 = 0.0972552$$



Splitter “Best”

$$\begin{aligned}
 \text{Info } a_{13}(D) &= \frac{34}{38} I(17, 17) + \frac{4}{36} I(4, 0) \\
 &= \frac{34}{38} \left[-\frac{17}{34} \log_2 \frac{17}{34} + \left(-\frac{17}{34} \log_2 \frac{17}{34} \right) \right] + \frac{4}{36} \left[-\frac{4}{4} \log_2 \frac{4}{4} + \left(-\frac{0}{4} \log_2 \frac{0}{4} \right) \right] \\
 &= \frac{34}{38} \times (0.5 + 0.5) \\
 &= 0.894737 //
 \end{aligned}$$

มูลค่าใด?



Splitter “Best”

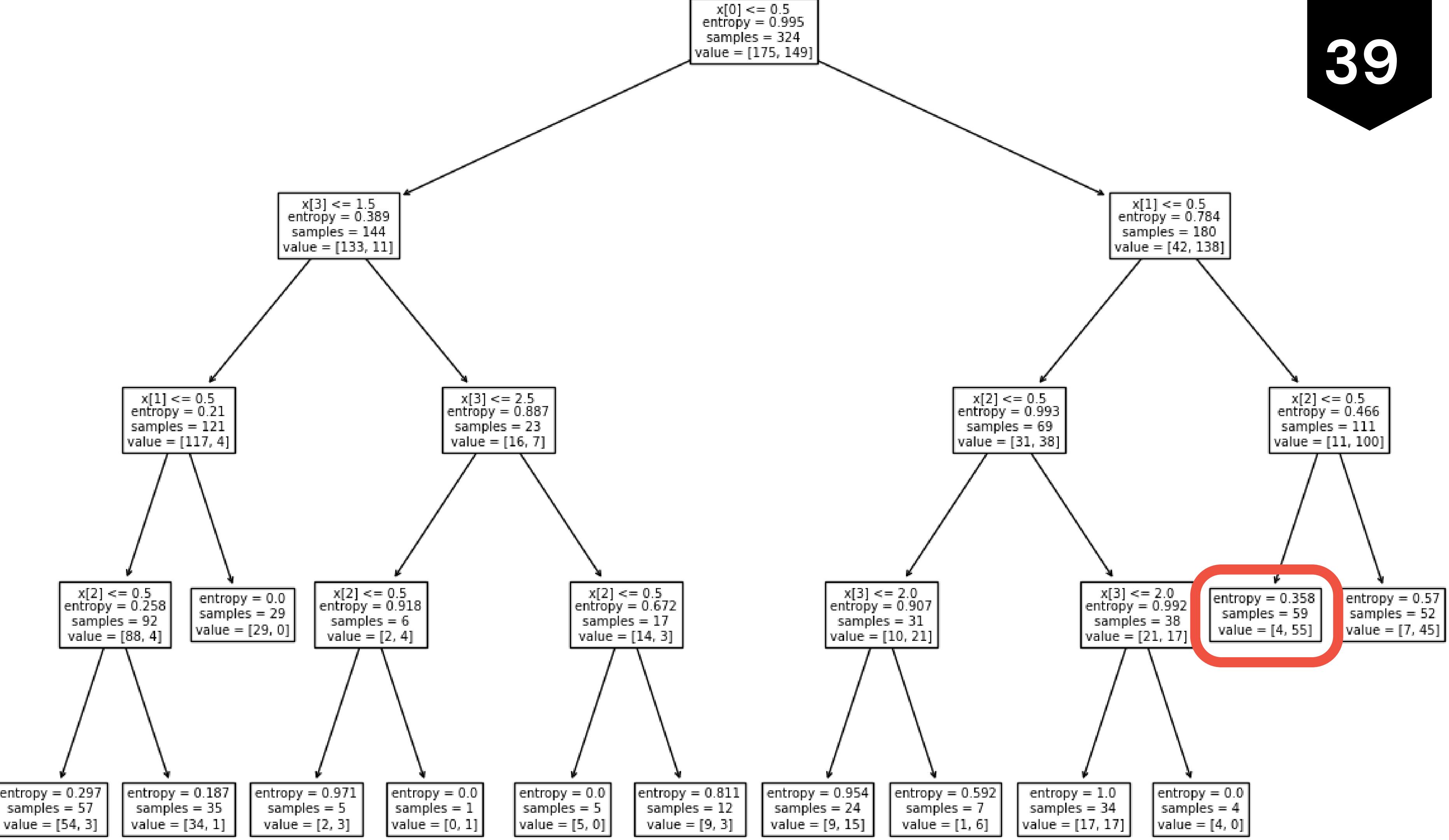
)
 $\text{กรณี } a_9 = 1, a_{10} = 1, a_{12} = 0$
 $\text{หรือ } a_9 \geq 0.5, a_{10} \geq 0.5, a_{12} \leq 0.5$

$$\text{Info}(D) = I(4, 55)$$

$$\begin{aligned}
 &= \left[-\frac{4}{59} \log_2 \frac{4}{59} \right] + \left[-\frac{55}{59} \log_2 \frac{55}{59} \right] \\
 &= 0.26323 + 0.0944167 \\
 &= 0.3576467,
 \end{aligned}$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
1	4	55	59
Grand Total	4	55	59

ไม่มีการแบ่งต่อ เพราะค่าที่เป็นไปได้ค่าเดียวคือ 1



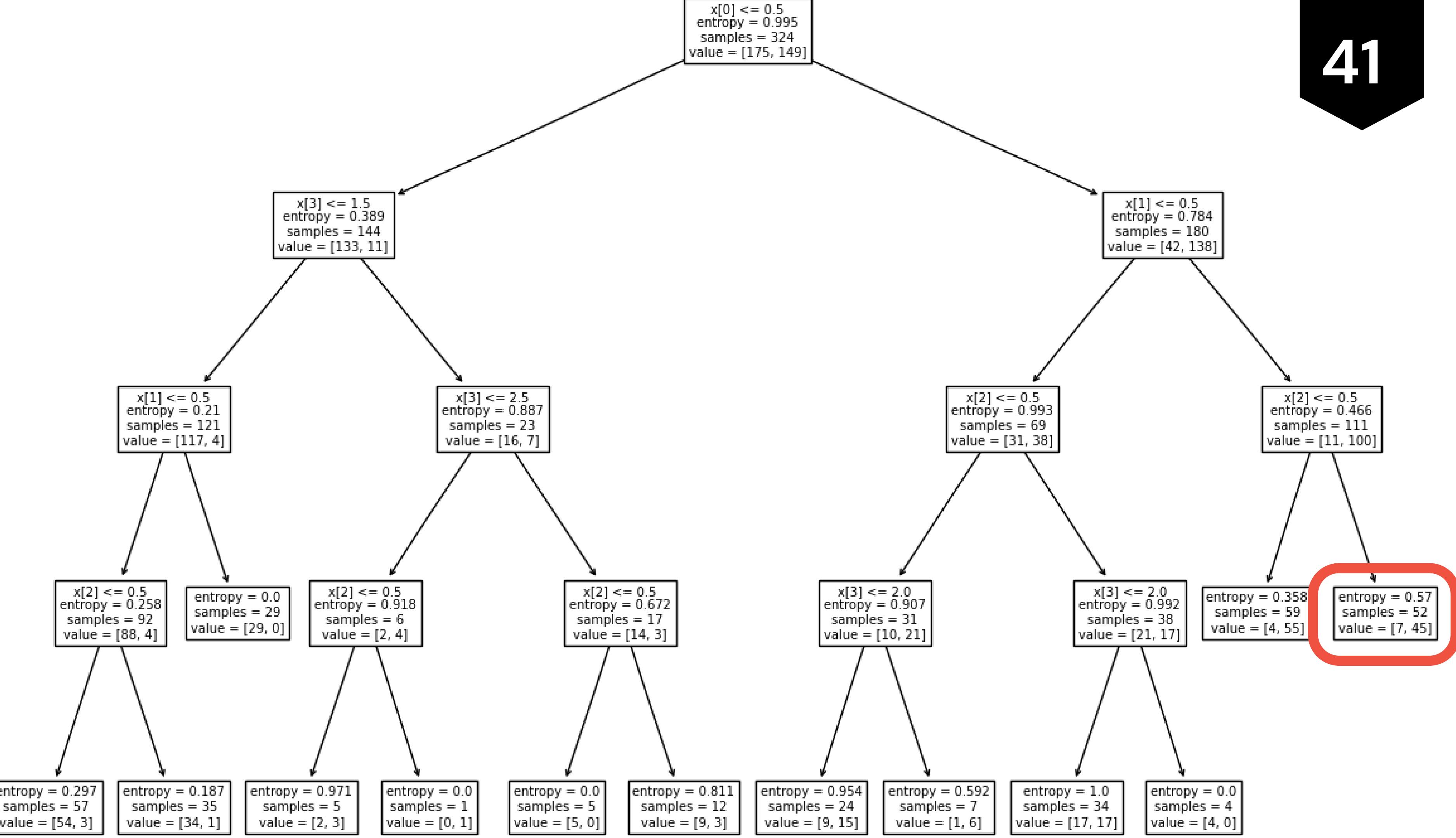
Splitter “Best”

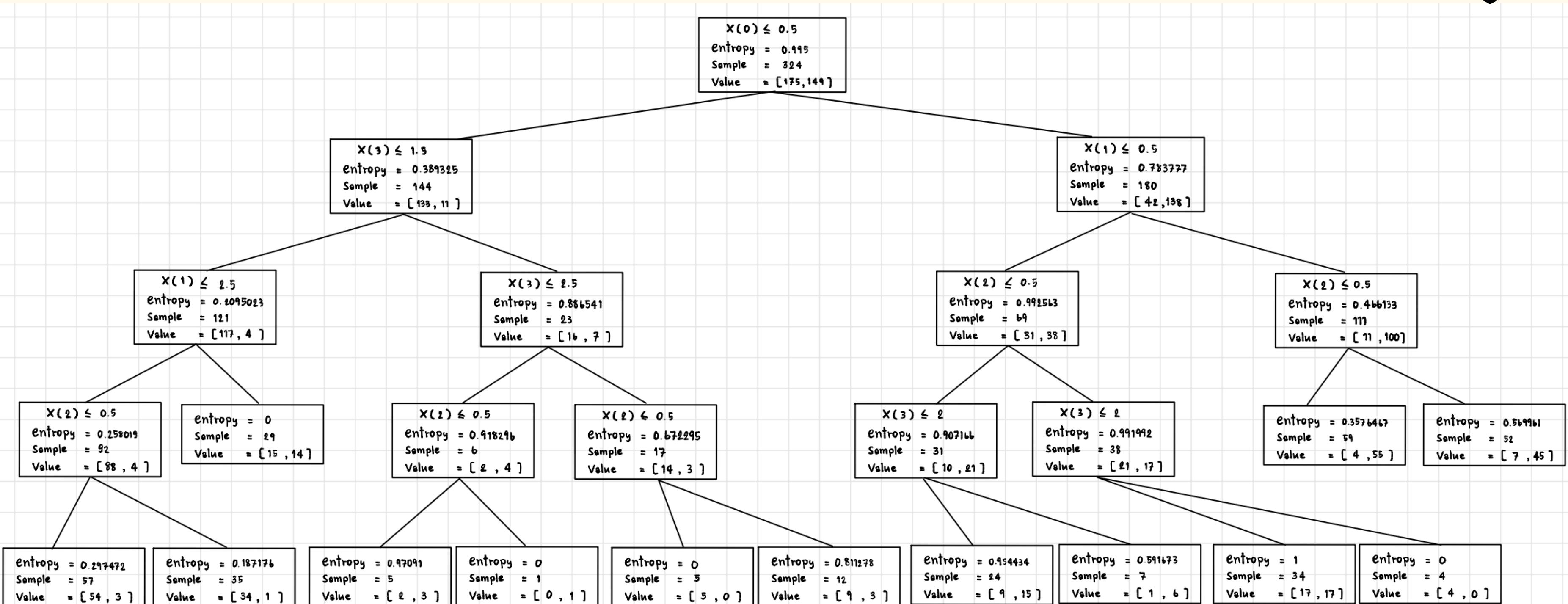
กรณี $\theta_9 = 1, \theta_{10} = 1, \theta_{12} = 1$
หรือ $\theta_9 \geq 0.5, \theta_{10} \geq 0.5, \theta_{12} \geq 0.5$

$$\begin{aligned} \text{Info}(D) &= I(7, 45) \\ &= \left[-\frac{7}{52} \log_2 \frac{7}{52} \right] + \left[-\frac{45}{52} \log_2 \frac{45}{52} \right] \\ &= 0.569961 \end{aligned}$$

Count of Column Labels			
Row La	0	1	Grand Total
1	7	45	52
Grand Tot	7	45	52

ไม่มีการแบ่งต่อ เพราะค่าที่เป็นไปได้ค่าเดียวคือ 1



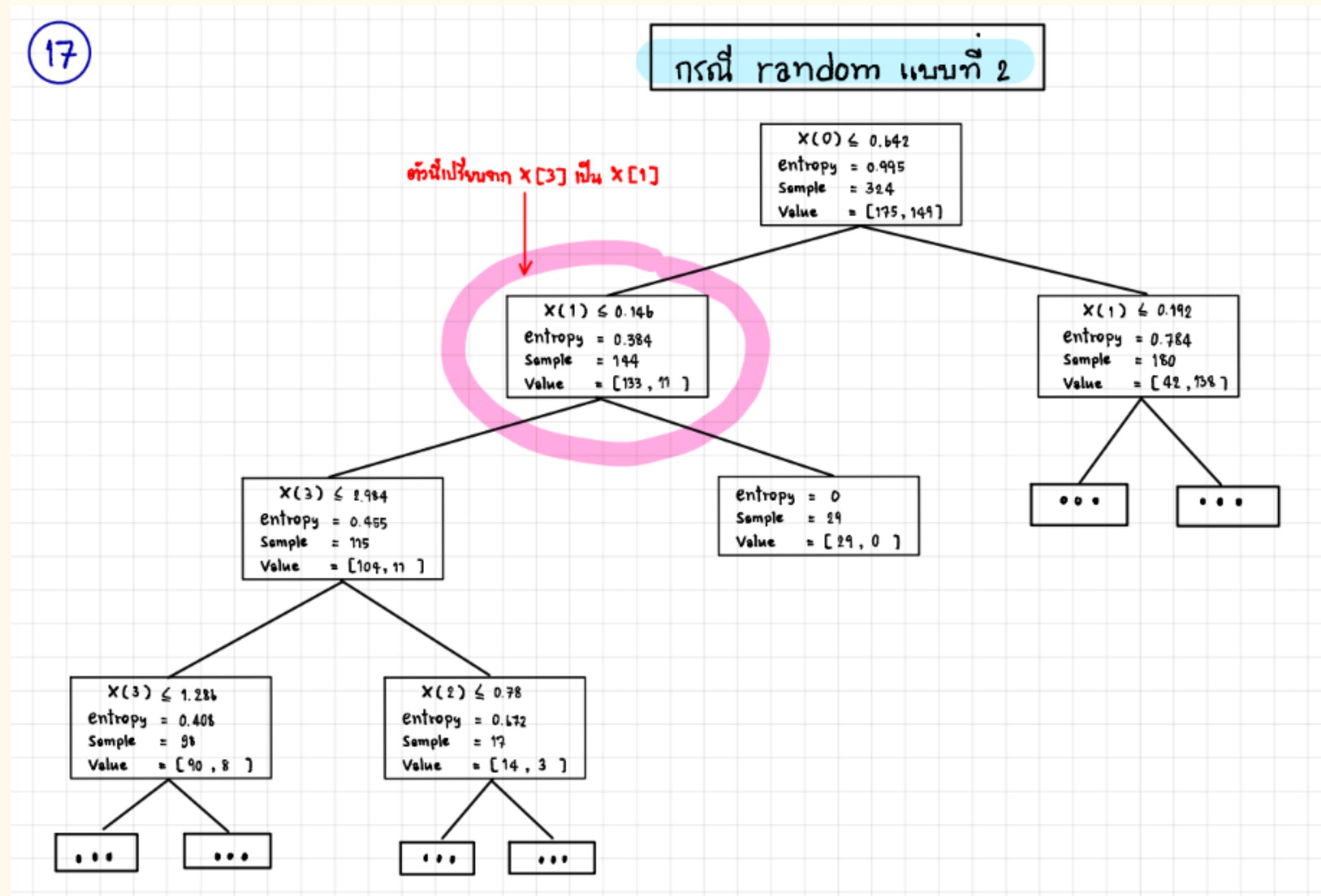


ความแม่นยำ = 83.88 %

Splitter “Random”



Splitter “Random”



Splitter “Random”

random หั้นๆ ก็เป็นไปได้

$$\begin{aligned} \text{กรณี } \theta_9 &\leq 0.299 \text{ และ } \theta_{10} \leq 0.744 \\ \text{หรือ } \theta_9 &= 0 \text{ และ } \theta_{10} = 0 \end{aligned}$$

$$Info(D) = I(104, 11)$$

$$\begin{aligned} &= \left[-\frac{104}{115} \log_2 \frac{104}{115} \right] + \left[-\frac{11}{115} \log_2 \frac{11}{115} \right] \\ &= 0.131176 + 0.323884 \\ &= 0.45506 // \end{aligned}$$

$$Gain(\theta_{12}) = 0.45506 - 0.45463 = 0.00043$$

$$\begin{aligned} Gain(\theta_{13}) &\rightarrow = 0.45506 - 0.383726 = 0.0678 \rightarrow \text{กรณี } (1, (2, 3)) \\ &\downarrow \rightarrow = 0.45506 - 0.4469897 = 0.0080703 \rightarrow \text{กรณี } ((1, 2), 3) \end{aligned}$$

Splitter “Random”

$$\begin{aligned}
 \text{Info } a_{12}(D) &= \frac{67}{115} I(61, b) + \frac{48}{115} I(43, 5) \\
 &= \frac{67}{115} \left[-\frac{61}{67} \log_2 \frac{61}{67} + \left(-\frac{6}{67} \log_2 \frac{6}{67} \right) \right] + \frac{48}{115} \left[-\frac{43}{48} \log_2 \frac{43}{48} + \left(-\frac{5}{48} \log_2 \frac{5}{48} \right) \right] \\
 &= 0.25342 + 0.20121 \\
 &= 0.45463 //
 \end{aligned}$$

Count of a16 Row Labels	Column Labels		Grand Total
	0	1	
0	61	6	67
1	43	5	48
Grand Total	104	11	115

Count of a16 Column Labels Row Labels	0		1 Grand Total
	0	1	
1	88	4	92
2	2	4	6
3	14	3	17
Grand Total	104	11	115

$$\begin{aligned}
 \text{Info } a_{13}(D) &= \frac{92}{115} I(88, 4) + \frac{23}{115} I(16, 7) \\
 &\downarrow \text{กรณี } ((1, (2, 3)) \\
 &= \frac{92}{115} \left[-\frac{88}{92} \log_2 \frac{88}{92} + \left(-\frac{4}{92} \log_2 \frac{4}{92} \right) \right] + \frac{23}{115} \left[-\frac{16}{23} \log_2 \frac{16}{23} + \left(-\frac{7}{23} \log_2 \frac{7}{23} \right) \right] \\
 &= 0.206415 + 0.177311 \\
 &= 0.383726 //
 \end{aligned}$$

$$\begin{aligned}
 \text{Info } a_{13}(D) &= \frac{98}{115} I(90, 8) + \frac{17}{115} I(14, 3) \\
 &\downarrow \text{กรณี } ((1, 2), 3) \\
 &= \frac{98}{115} \left[-\frac{90}{98} \log_2 \frac{90}{98} + \left(-\frac{8}{98} \log_2 \frac{8}{98} \right) \right] + \frac{17}{115} \left[-\frac{14}{17} \log_2 \frac{14}{17} + \left(-\frac{3}{17} \log_2 \frac{3}{17} \right) \right] \\
 &= 0.347607 + 0.0993827 \\
 &= 0.4469897
 \end{aligned}$$

Splitter “Random”

19

กรณี $a_9 = 0, a_{10} = 0, a_{13} = 1$ หรือ 2
แล้ว $a_9 \leq 0.74, a_{10} \leq 0.072, a_{13} \leq 2.984$

$$\begin{aligned}
 \text{Info}(D) &= I(90, 8) \\
 &= \left[-\frac{90}{98} \log_2 \frac{90}{98} \right] + \left[-\frac{8}{98} \log_2 \frac{8}{98} \right] \\
 &= 0.112828 + 0.295078 \\
 &= 0.407906 //
 \end{aligned}$$

Count of a16		Column Labels	
		0	1
Row Labels	Grand Total		
	0	56	6
1	34	2	36
Grand Total	90	8	98

$$\begin{aligned}
 \text{Info } a_{12}(D) &= \frac{62}{98} I(56, 6) + \frac{36}{98} I(34, 2) \\
 &= \frac{62}{98} \left[-\frac{56}{62} \log_2 \frac{56}{62} + \left(-\frac{6}{62} \log_2 \frac{6}{62} \right) \right] + \frac{36}{98} \left[-\frac{34}{36} \log_2 \frac{34}{36} + \left(-\frac{2}{36} \log_2 \frac{2}{36} \right) \right] \\
 &= 0.290189 + 0.11371 \\
 &= 0.403899
 \end{aligned}$$

Splitter “Random”

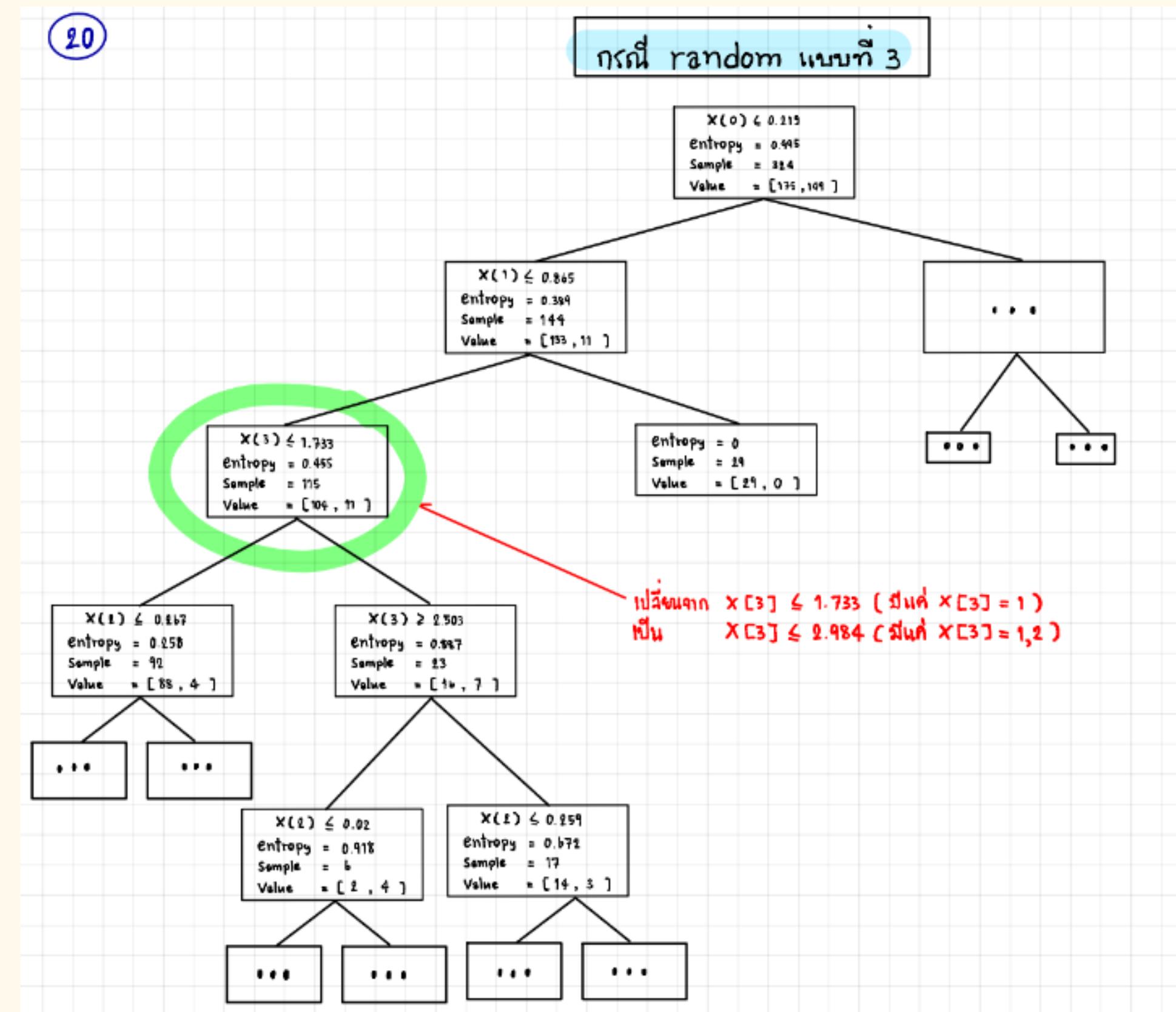
$$\begin{aligned}
 \text{Info } a_{13}(D) &= \frac{92}{98} I(88, 4) + \frac{b}{98} I(2, 4) \\
 \downarrow \\
 \text{Gain } (a_{13}) &= \frac{92}{98} \left[-\frac{88}{92} \log_2 \frac{88}{92} + \left(-\frac{4}{92} \log_2 \frac{4}{92} \right) \right] + \frac{b}{98} \left[-\frac{2}{b} \log_2 \frac{2}{b} + \left(-\frac{4}{b} \log_2 \frac{4}{b} \right) \right] \\
 &= 0.407906 + 0.0562222 \\
 &= 0.407906
 \end{aligned}$$

		Count of a16		Column Labels		
		Row Labels		0	1	Grand Total
	1			88	4	92
	2			2	4	6
	Grand Total			90	8	98

$$\text{Gain } (a_{12}) = 0.407906 - 0.403899 = 0.004007$$

$$\text{Gain } (a_{13}) = 0.407906 - 0.298442 = 0.109464$$

Splitter “Random”



Splitter “Random”

กรณี $a_9 = 0, a_{10} = 0, a_{13} = 2 \leq 3$
 ให้ $a_9 \leq 0.74, a_{10} \leq 0.072, a_{13} \geq 1.389$

$$\begin{aligned}
 \text{Info}(D) &= I(16, 7) \\
 &= \left[-\frac{16}{23} \log_2 \frac{16}{23} \right] + \left[-\frac{7}{23} \log_2 \frac{7}{23} \right] \\
 &= 0.364217 + 0.522324 \\
 &= 0.886541,
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(a_{12}) &= 0.886541 - 0.886492 = 0.000049 \\
 \text{Gain}(a_{13}) &= 0.886541 - 0.736469 = 0.150072
 \end{aligned}$$

Splitter “Random”

$$\begin{aligned}
 \text{Info } a_{13}(D) &= \frac{6}{23} I(2, 4) + \frac{17}{23} I(14, 3) \\
 &= \frac{6}{23} \left[-\frac{2}{6} \log_2 \frac{2}{6} + \left(-\frac{4}{6} \log_2 \frac{4}{6} \right) \right] + \frac{17}{23} \left[-\frac{14}{17} \log_2 \frac{14}{17} + \left(-\frac{3}{17} \log_2 \frac{3}{17} \right) \right] \\
 &= 0.239555 + 0.496914 \\
 &= 0.736469 //
 \end{aligned}$$

Count of a16		Column Labels		
Row Labels		0	1	Grand Total
0		7	3	10
1		9	4	13
Grand Total		16	7	23

Splitter “Random”

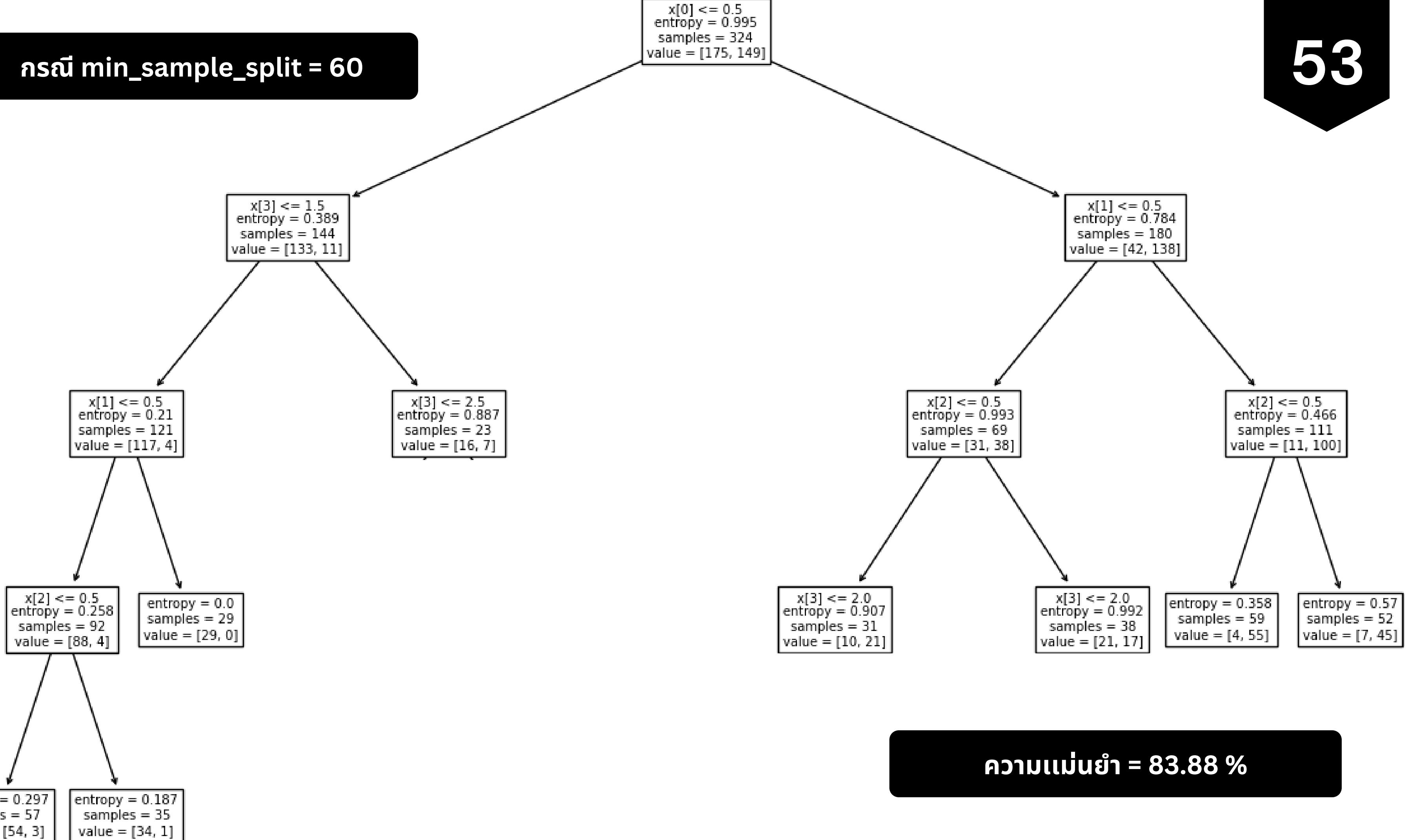
$$\begin{aligned}
 \text{Info}_{\text{a}_{12}}(\text{D}) &= \frac{10}{23} I(7, 3) + \frac{13}{23} I(9, 4) \\
 &= \frac{10}{23} \left[-\frac{7}{10} \log_2 \frac{7}{10} + \left(-\frac{3}{10} \log_2 \frac{3}{10} \right) \right] + \frac{13}{23} \left[-\frac{9}{13} \log_2 \frac{9}{13} + \left(-\frac{4}{13} \log_2 \frac{4}{13} \right) \right] \\
 &= 0.38317 + 0.503322 \\
 &= 0.886492 //
 \end{aligned}$$

Count of a Column Labels			
Row La		0	1 Grand Total
2		2	4
3		14	3
Grand Tot		16	7
			23

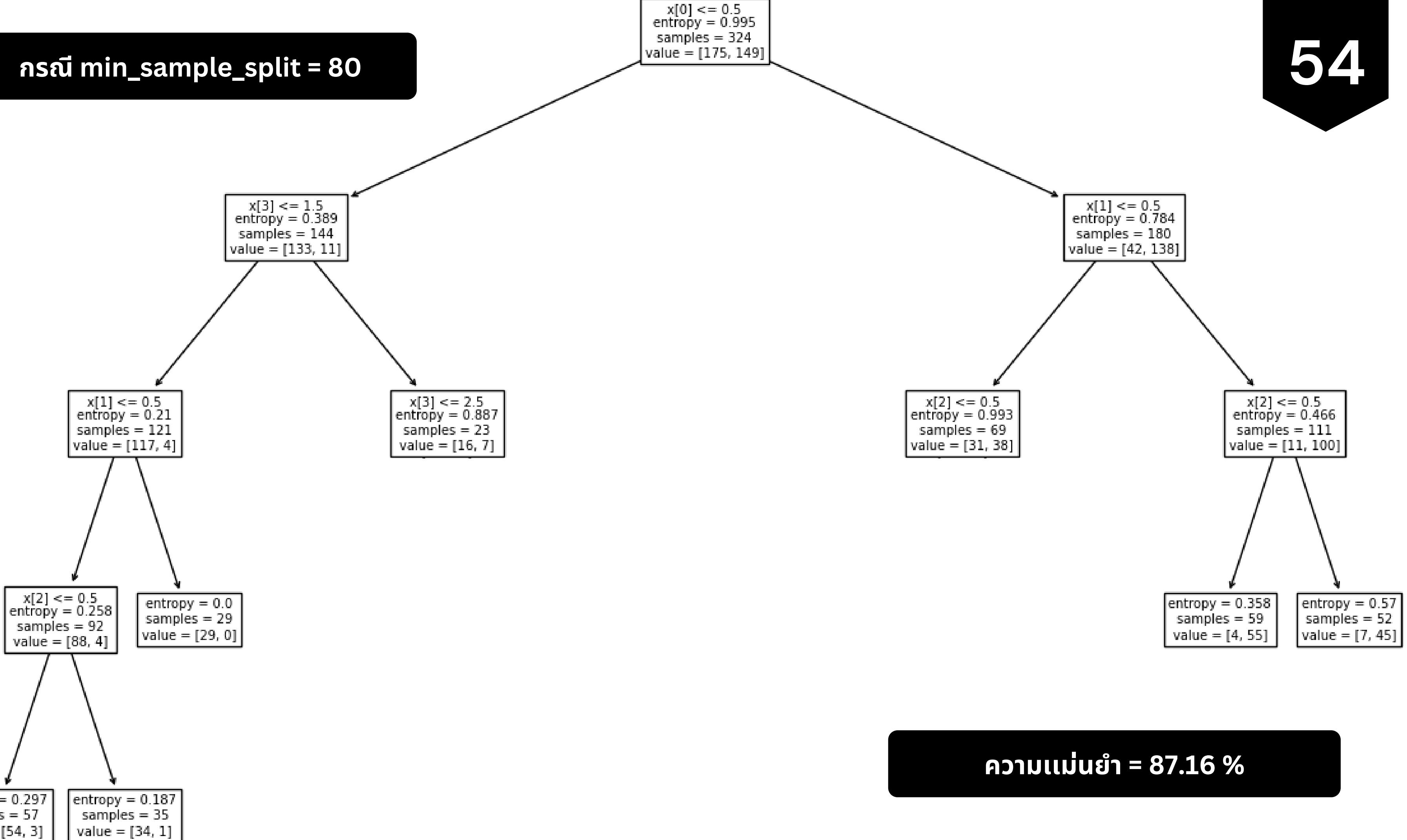
Min Samples Split



กรณี min_sample_split = 60

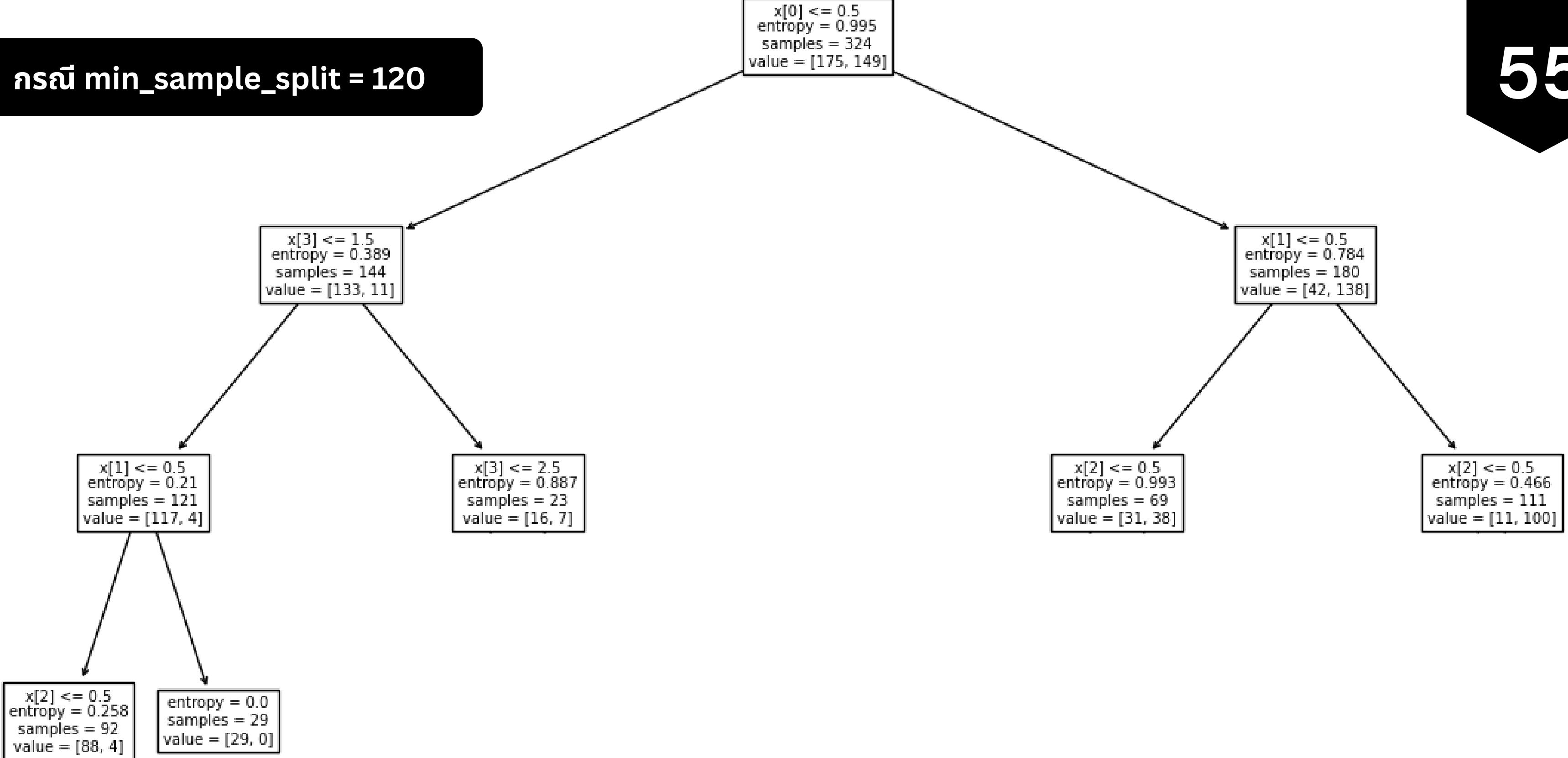


กรณี min_sample_split = 80



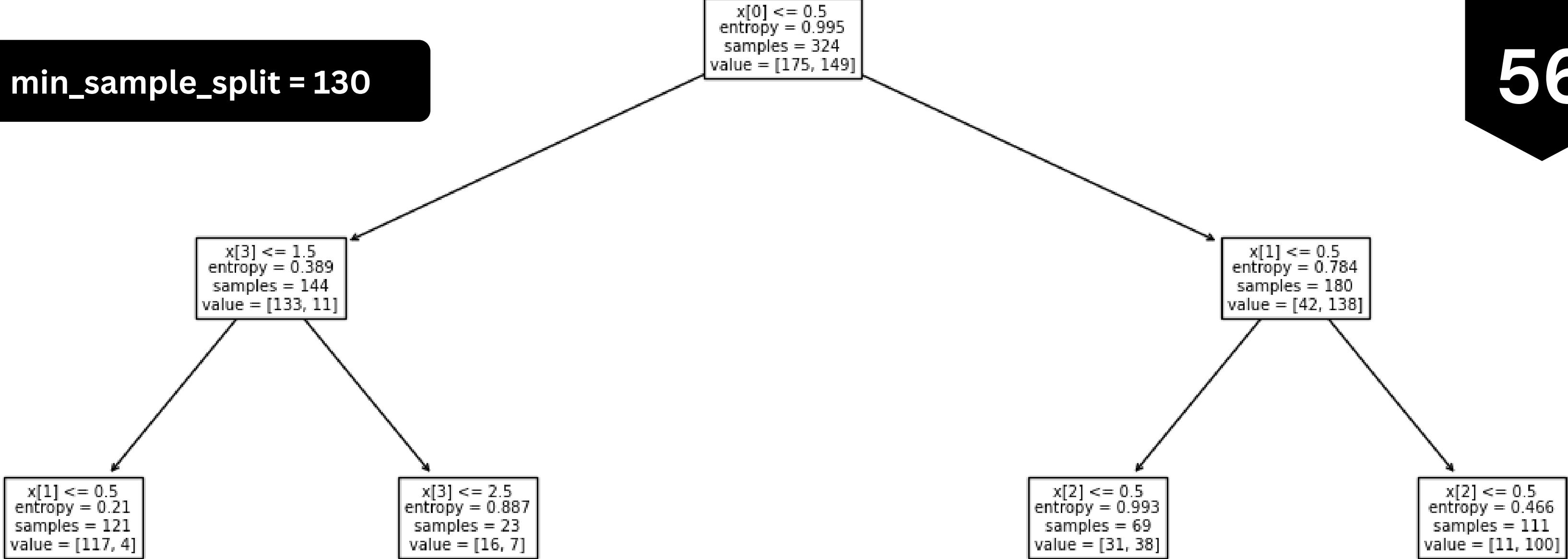
ความแม่นยำ = 87.16 %

กรณี min_sample_split = 120



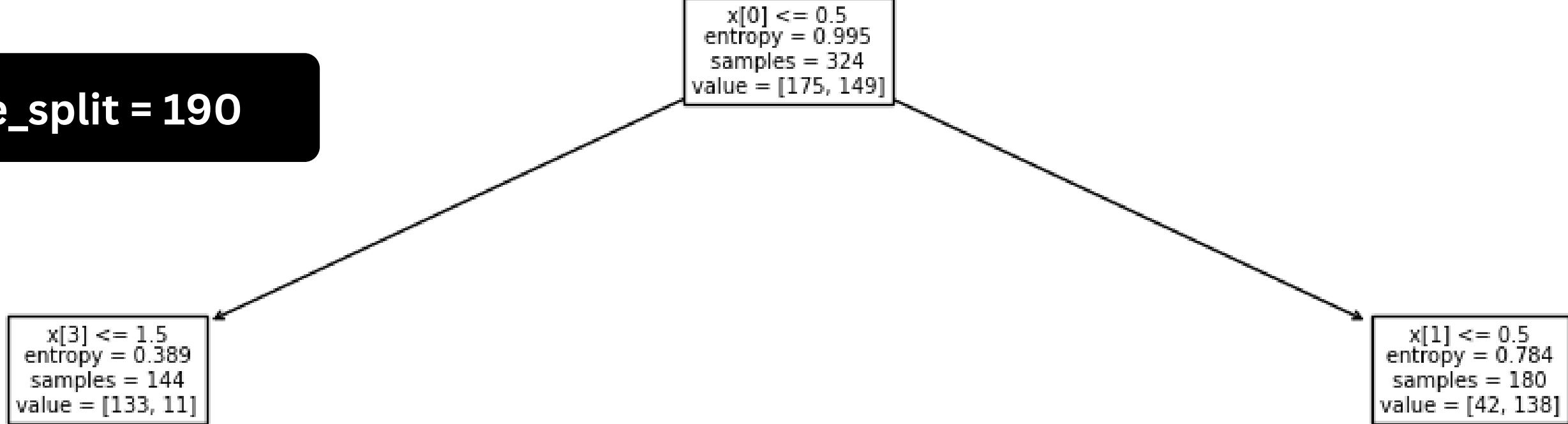
ความแม่นยำ = 87.16 %

กรณี min_sample_split = 130



ความแม่นยำ = 87.16 %

กรณี `min_sample_split = 190`



ความแม่นยำ = 87.16 %

THANK YOU

