HOMEWORK 4

DECISION TREE CLASSIFIER

Presented by: NMN & หนูน้อยหอยนางรม

MEMBER

นางสาวเพ็ญพิชชา วรรณ์ชูมาตร์
643020063-4

นายภัทรธร ก้อนมณี
643020513-9

นายรัชชานนท์ ทิพย์พิมานพร
643020515-5

นายวุฒิชัย คำนา
643020521-0

นางสาววริศรา ปันลา
643020519-7

นางสาวอนัญญา พูลสวัสดิ์
643020526-0

นางสาววิลันดา ทาระมาตย์
643021271-2

Parameters

Min_Samples_Split

เป็นพารามิเตอร์ที่ควบคุมจำนวนตัวอย่างขั้นต่ำที่ต้องอยู่ในโหนดใดโหนดหนึ่งก่อนที่จะแบ่งโหนดนั้นต่อไป (โหนดจะไม่ถูกแบ่งออกหากมีจำนวนตัวอย่างน้อยกว่า)

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 treadshold ไว้ค่าหนึ่งที่ไว้ใช้แบ่งซึ่งจะเป็นค่ากลางเสมอ เช่น 0.5 แล้วทำการแบ่งข้อมูล

Random

จะมีการสุ่ม treadshold ในทุกๆ feature ก่อน แล้วค่อยหาค่า entropy หรือ gini ที่ดีที่สุด (ถ้าเป็นentropy ต้องการค่า gain สูงมากสุด) แล้วทำการแบ่งข้อมูล

Splitter “Best”

Define

x0 = a9

x1 = a10

x2 = a12

x3 = a13

y = a16

Splitter “Best”

Define

x0 = a9

x1 = a10

x2 = a12

x3 = a13

y = a16

Entropy

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

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

$$= 0.479975 + 0.515375$$

$$= 0.99535$$

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

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

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

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

$$\text{Gain}(a_{16}) \rightarrow 1,(2,3) = 0.99535 - 0.993281 = 0.002069$$

$$(1,2),3 = 0.99535 - 0.989944 = 0.005406$$

Splitter “Best”

Define

x0 = a9

x1 = a10

x2 = a12

x3 = a13

y = a16

Entropy

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

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

$$= 0.479975 + 0.515375$$

$$= 0.99535$$

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

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

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

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

$$\text{Gain}(a_{15}) \rightarrow 1,(2,3) = 0.99535 - 0.993281 = 0.002069$$

$$(1,2),3 = 0.99535 - 0.989944 = 0.005406$$

$$X(0) \leq 0.5$$

$$\text{entropy} = 0.995$$

$$\text{Sample} = 324$$

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

Splitter “Best”

a9

$$\text{Info}_{a_9}(D) = \frac{144}{324} I(11,133) + \frac{180}{324} I(138,42)$$

$$= \frac{144}{324}\left[-\frac{11}{144}\log_2\frac{11}{144} + \left(-\frac{133}{144}\log_2\frac{133}{144}\right)\right] + \frac{180}{324}\left[-\frac{138}{180}\log_2\frac{138}{180} + \left(-\frac{42}{180}\log_2\frac{42}{180}\right)\right]$$

$$= 0.173033 + 0.435421$$

$$= 0.608454 //$$

Count of a16	Column La		
Row Labels	0	1	Grand Total
0	133	11	144
1	42	138	180
Grand Total	175	149	324

a10

$$\text{Info}_{a_{10}}(D) = \frac{184}{324} I(135,49) + \frac{140}{324} I(40,100)$$

$$= \frac{184}{324}\left[-\frac{135}{184}\log_2\frac{135}{184} + \left(-\frac{49}{184}\log_2\frac{49}{184}\right)\right] + \frac{140}{324}\left[-\frac{40}{140}\log_2\frac{40}{140} + \left(-\frac{100}{140}\log_2\frac{100}{140}\right)\right]$$

$$= 0.47483 + 0.372953$$

$$= 0.84778 //$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
0	135	49	184
1	40	100	140
Grand Total	175	149	324

a12

$$\text{Info}_{a_{12}}(D) = \frac{172}{324} I(90,82) + \frac{152}{324} I(85,67)$$

$$= \frac{172}{324}\left[-\frac{90}{172}\log_2\frac{90}{172} + \left(-\frac{82}{172}\log_2\frac{82}{172}\right)\right] + \frac{152}{324}\left[-\frac{85}{152}\log_2\frac{85}{152} + \left(-\frac{67}{152}\log_2\frac{67}{152}\right)\right]$$

$$= 0.530036 + 0.464375$$

$$= 0.99422 //$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
0	90	82	172
1	85	67	152
Grand Total	175	149	324

Splitter “Best”

a13

$$\text{Info}_{a_{13}}(D) = \frac{290}{324} I(154,136) + \frac{6}{324} I(2,4) + \frac{28}{324} I(19,9)$$

$$= \frac{290}{324}\left[-\frac{154}{290}\log_2\frac{154}{290} + \left(-\frac{136}{290}\log_2\frac{136}{290}\right)\right] + \frac{6}{324}\left[-\frac{2}{6}\log_2\frac{2}{6} + \left(-\frac{4}{6}\log_2\frac{4}{6}\right)\right]$$

$$+ \frac{28}{324}\left[-\frac{19}{28}\log_2\frac{19}{28} + \left(-\frac{9}{28}\log_2\frac{9}{28}\right)\right]$$

$$= 0.892573 + 0.0170055 + 0.078290$$

$$= 0.98787 //$$

a13
1, (2,3)

$$\text{Info}_{a_{13}}(D) = \frac{290}{324} I(154,136) + \frac{34}{324} I(21,13)$$

ที่แบ่งเป็น 1,(2,3)

$$= \frac{290}{324}\left[-\frac{154}{290}\log_2\frac{154}{290} + \left(-\frac{136}{290}\log_2\frac{136}{290}\right)\right] + \frac{34}{324}\left[-\frac{21}{34}\log_2\frac{21}{34} + \left(-\frac{13}{34}\log_2\frac{13}{34}\right)\right]$$

$$= 0.892573 + 0.100706$$

$$= 0.993281 //$$

Count of a16	Column La		
Row Labels	0	1	Grand Total
1	154	136	290
2	2	4	6
3	19	9	28
Grand Total	175	149	324

a13
(1,2), 3

$$\text{Info}_{a_{13}}(D) = \frac{296}{324} I(156,140) + \frac{28}{324} I(19,9)$$

แบ่งเป็น 1,(2,3)

$$= \frac{296}{324}\left[-\frac{156}{296}\log_2\frac{156}{296} + \left(-\frac{140}{296}\log_2\frac{140}{296}\right)\right] + \frac{28}{324}\left[-\frac{19}{28}\log_2\frac{19}{28} + \left(-\frac{9}{28}\log_2\frac{9}{28}\right)\right]$$

$$= 0.911654 + 0.0782901$$

$$= 0.969944 //$$

Splitter “Best”

กรณี $a_9 = 0$
หรือ $a_9 \leq 0.5$

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

$$= \left[-\frac{133}{144} \log_2 \frac{133}{144} \right] + \left[-\frac{11}{144} \log_2 \frac{11}{144} \right]$$

$$= 0.28344 + 0.105885$$

$$= 0.389325$$

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

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

$$\text{Gain}(a_{13}) \begin{cases} = 0.389325 - 0.31764 = \boxed{0.071685} \\ = 0.389325 - 0.378532 = 0.010793 \end{cases}$$

→ กรณี a_{13} ที่แบ่งเป็นกรณี 1, (2,3)

Splitter “Best”

$a_9 = 0$

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

$$= \left[-\frac{133}{144}\log_2\frac{133}{144}\right] + \left[-\frac{11}{144}\log_2\frac{11}{144}\right]$$

$$= 0.28344 + 0.105885$$

$$= 0.389325$$

$X(0) \leq 0.5$
entropy = 0.995
Sample = 324
Value = [175, 149]

$X(3) \leq 1.5$
entropy = 0.389325
Sample = 144
Value = [133, 11]

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

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

$$\text{Gain}(a_{13}) \rightarrow = 0.389325 - 0.31764 = \boxed{0.071685}$$

$$= 0.389325 - 0.378532 = 0.010793$$

→ กรณี a_{13} ที่แบ่งเป็นกรณี 1, (2,3)

Splitter “Best”

$$Info_{a_{10}}(D) = \frac{115}{144} I(104,11) + \frac{29}{144} I(29,0)$$

$$= \frac{115}{144}\left[-\frac{104}{115}\log_2\frac{104}{115} + \left(-\frac{11}{115}\log_2\frac{11}{115}\right)\right] + \frac{29}{144}\left[-\frac{29}{29}\log_2\frac{29}{29} + \left(-\frac{0}{29}\log_2\frac{0}{29}\right)\right]$$

0 หาค่าไม่ได้

$$= 0.363416$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
0	104	11	115
1	29		29
Grand Total	133	11	144

Count of a16	Column La	s	
Row Labels	0	1	Grand Total
0	76	6	82
1	57	5	62
Grand Total	133	11	144

$$Info_{a_{12}}(D) = \frac{82}{144} I(76,6) + \frac{62}{144} I(57,5)$$

$$= \frac{82}{144}\left[-\frac{76}{82}\log_2\frac{76}{82} + \left(-\frac{6}{82}\log_2\frac{6}{82}\right)\right] + \frac{62}{144}\left[-\frac{57}{62}\log_2\frac{57}{62} + \left(-\frac{5}{62}\log_2\frac{5}{62}\right)\right]$$

$$= 0.215048 + 0.174149$$

$$= 0.389197$$

$Info_{a_{13}}(D)$ แบ่งเป็น 1,(2,3)

$$= \frac{121}{144} I(117,4) + \frac{23}{144} I(16,7)$$

$$= \frac{121}{144}\left[-\frac{117}{121}\log_2\frac{117}{121} + \left(-\frac{4}{121}\log_2\frac{4}{121}\right)\right] + \frac{23}{144}\left[-\frac{16}{23}\log_2\frac{16}{23} + \left(-\frac{7}{23}\log_2\frac{7}{23}\right)\right]$$

$$= 0.17604 + 0.1416$$

$$= 0.31764$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
1	117	4	121
2	2	4	6
3	14	3	17
Grand Total	133	11	144

$Info_{a_{13}}(D)$ แบ่งเป็น (1,2),3

$$= \frac{127}{144} I(119,8) + \frac{17}{144} I(14,3)$$

$$= \frac{127}{144}\left[-\frac{119}{127}\log_2\frac{119}{127} + \left(-\frac{8}{127}\log_2\frac{8}{127}\right)\right] + \frac{17}{144}\left[-\frac{14}{17}\log_2\frac{14}{17} + \left(-\frac{3}{17}\log_2\frac{3}{17}\right)\right]$$

$$= 0.299164 + 0.0793682$$

$$= 0.378532$$

Splitter “Best”

กรณี $a_9 = 1$
หรือ $a_9 \geq 0.5$

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

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

$$= 0.489892 + 0.293885$$

$$= 0.783777 \text{ บิต}$$

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

$$\text{Gain}(a_{12}) = 0.783777 - 0.759011 = 0.0024766$$

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

Splitter “Best”

กรณี $a_9 = 1$
หรือ $a_9 \geq 0.5$

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

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

$$= 0.489892 + 0.293885$$

$$= 0.783777$$

$X(0) \leq 0.5$
Entropy = 0.995
Sample = 324
Value = [175, 149]

$X(3) \leq 1.5$
Entropy = 0.389325
Sample = 144
Value = [133, 11]

$X(1) \leq 0.5$
Entropy = 0.783777
Sample = 180
Value = [42, 138]

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

Splitter “Best”

$$\text{Info}_{a_{10}}(D) = \frac{69}{180} I(31,38) + \frac{111}{180} I(11,100)$$
$$= \frac{69}{180}\left[-\frac{31}{69}\log_2\frac{31}{69} + \left(-\frac{38}{69}\log_2\frac{38}{69}\right)\right] + \frac{111}{180}\left[-\frac{11}{111}\log_2\frac{11}{111} + \left(-\frac{100}{111}\log_2\frac{100}{111}\right)\right]$$
$$= 0.380482 + 0.287449$$
$$= 0.667931 //$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
0	31	38	69
1	11	100	111
Grand Total	42	138	180

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
0	14	76	90
1	28	62	90
Grand Total	42	138	180

$$\text{Info}_{a_{12}}(D) = \frac{90}{160} I(14,76) + \frac{90}{180} I(28,62)$$
$$= \frac{90}{180}\left[-\frac{14}{90}\log_2\frac{14}{90} + \left(-\frac{76}{90}\log_2\frac{76}{90}\right)\right] + \frac{90}{180}\left[-\frac{28}{90}\log_2\frac{28}{90} + \left(-\frac{62}{90}\log_2\frac{62}{90}\right)\right]$$
$$= 0.311785 + 0.447226$$
$$= 0.759011 //$$

$\text{Info}_{a_{13}}(D)$ แบ่งเป็น (1,3)

$$= \frac{169}{180} I(37,132) + \frac{11}{180} I(5,6)$$
$$= \frac{169}{180}\left[-\frac{37}{169}\log_2\frac{37}{169} + \left(-\frac{132}{169}\log_2\frac{132}{169}\right)\right] + \frac{11}{180}\left[-\frac{5}{11}\log_2\frac{5}{11} + \left(-\frac{6}{11}\log_2\frac{6}{11}\right)\right]$$
$$= 0.711882 + 0.0607463$$
$$= 0.7726283 //$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
1	37	132	169
3	5	6	11
Grand Total	42	138	180

Splitter “Best”

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

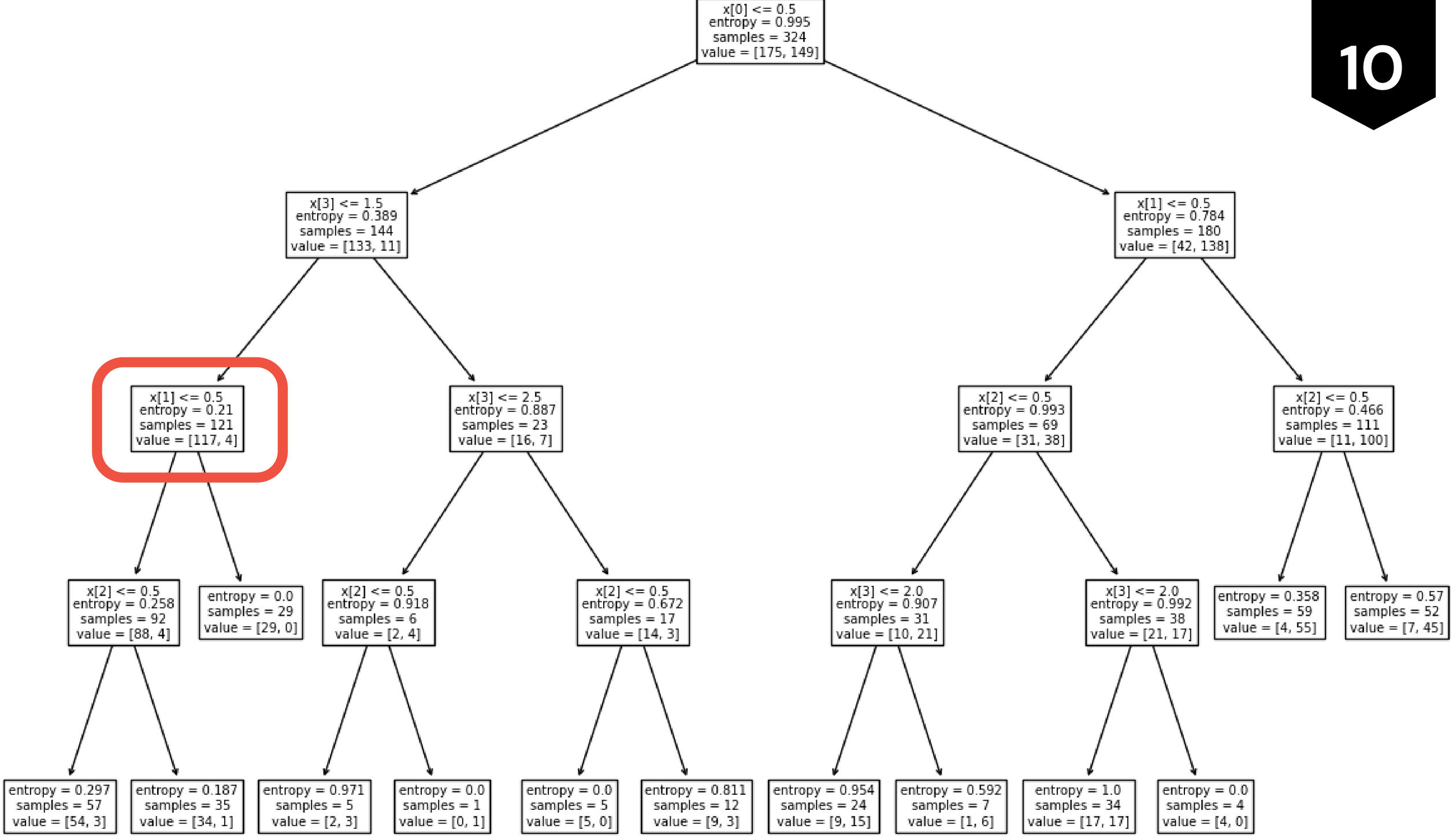
$$\text{Info}(D) = I(117,4)$$

$$= \left[-\frac{117}{121}\log_2\frac{117}{121}\right] + \left[-\frac{4}{121}\log_2\frac{4}{121}\right]$$

$$= 0.0468953 + 0.162607$$

$$= 0.2095023$$

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


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Best Samples Split

Count of a16	Column		
Row Labels	0	1	Grand Total
0	88	4	92
1	29		29
Grand Total	117	4	121

$$Info_{a10}(D) = \frac{92}{121} I(88,4) + \frac{29}{121} I(29,0)$$

$$= \frac{92}{121}\left[-\frac{88}{92}\log_2\frac{88}{92} + \left(-\frac{4}{92}\log_2\frac{4}{92}\right)\right] + \frac{29}{121}\left[-\frac{29}{29}\log_2\frac{29}{29} + \left(-\frac{0}{29}\log_2\frac{0}{29}\right)\right]$$

0 หารไม่ได้

$$= 0.19618 //$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
0	69	3	72
1	48	1	49
Grand Total	117	4	121

$$Info_{a12}(D) = \frac{72}{121} I(69,3) + \frac{49}{121} I(48,1)$$

$$= \frac{72}{121}\left[-\frac{69}{72}\log_2\frac{69}{72} + \left(-\frac{3}{72}\log_2\frac{3}{72}\right)\right] + \frac{49}{121}\left[-\frac{48}{49}\log_2\frac{48}{49} + \left(-\frac{1}{49}\log_2\frac{1}{49}\right)\right]$$

$$= 0.14869 + 0.0582032$$

$$= 0.2068932 //$$

Best Samples Split

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

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

$$= \left[-\frac{16}{23} \log_2 \frac{16}{23} \right] + \left[-\frac{7}{23} \log_2 \frac{7}{23} \right]$$

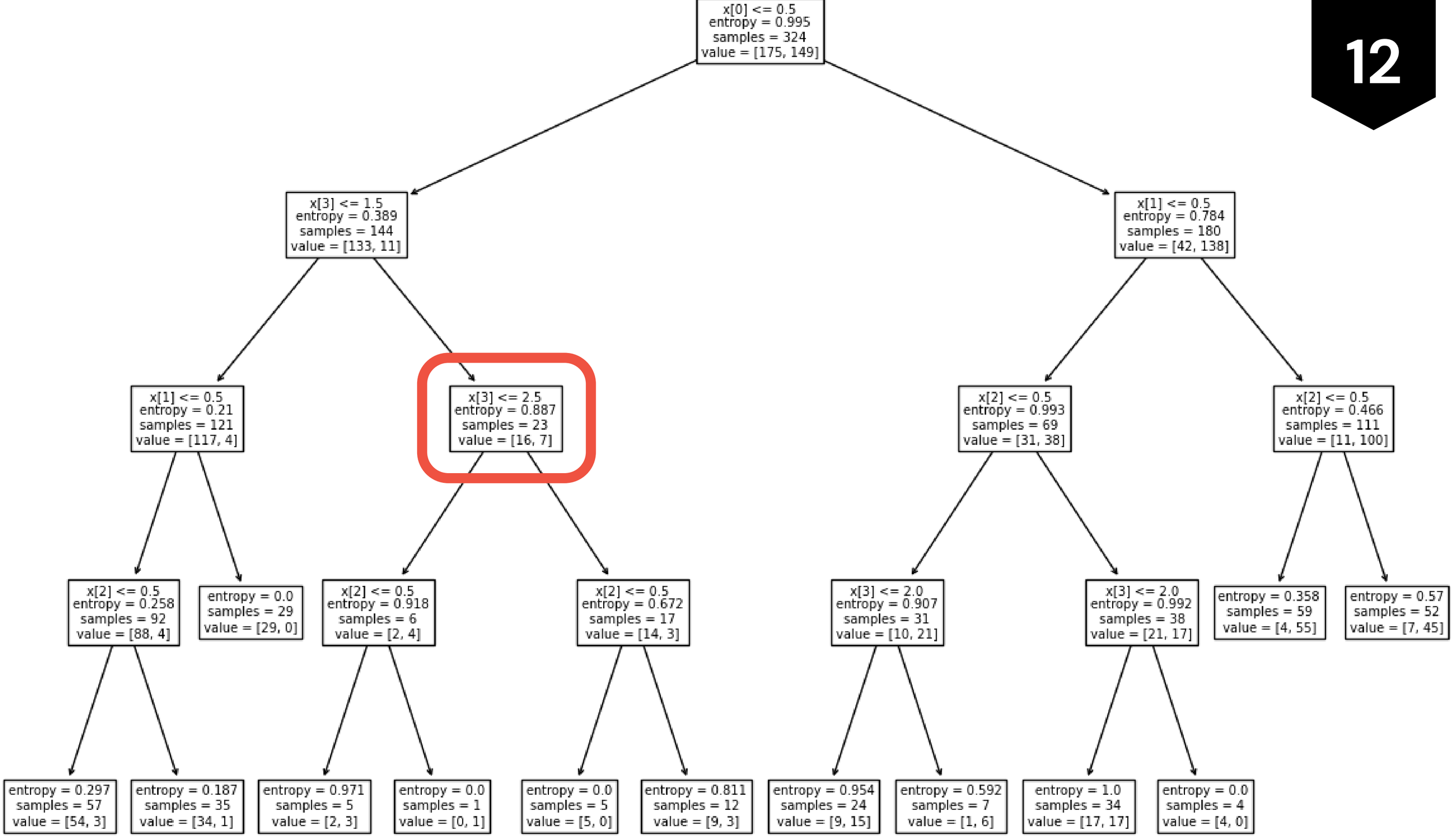
$$= 0.364217 + 0.522324$$

$$= 0.886541 \text{ //}$$

$$\text{Gain}(a_{10}) = 0.886541 - 0.886541 = 0$$

$$\text{Gain}(a_{12}) = 0.886541 - 0.88641 = 0.000051$$

$$\text{Gain}(a_{13}) = 0.886541 - 0.736469 = \boxed{0.150072}$$


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter "Best"

$$\text{Info}_{a_{10}}(D) = \frac{23}{23} I(16,7)$$
$$= \frac{23}{23}\left[-\frac{16}{23}\log_2\frac{16}{23} + \left(-\frac{7}{23}\log_2\frac{7}{23}\right)\right]$$
$$= 0.886541 //$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
0	16	7	23
Grand Total	16	7	23

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

$$\text{Info}_{a_{12}}(D) = \frac{10}{23} I(7,3) + \frac{13}{23} I(9,4)$$
$$= \frac{10}{23}\left[-\frac{7}{10}\log_2\frac{7}{10} + \left(-\frac{3}{10}\log_2\frac{3}{10}\right)\right] + \frac{13}{23}\left[-\frac{9}{13}\log_2\frac{9}{13} + \left(-\frac{4}{13}\log_2\frac{4}{13}\right)\right]$$
$$= 0.38317 + 0.503321$$
$$= 0.88649 //$$

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

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
2	2	4	6
3	14	3	17
Grand Total	16	7	23

Splitter “Best”

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

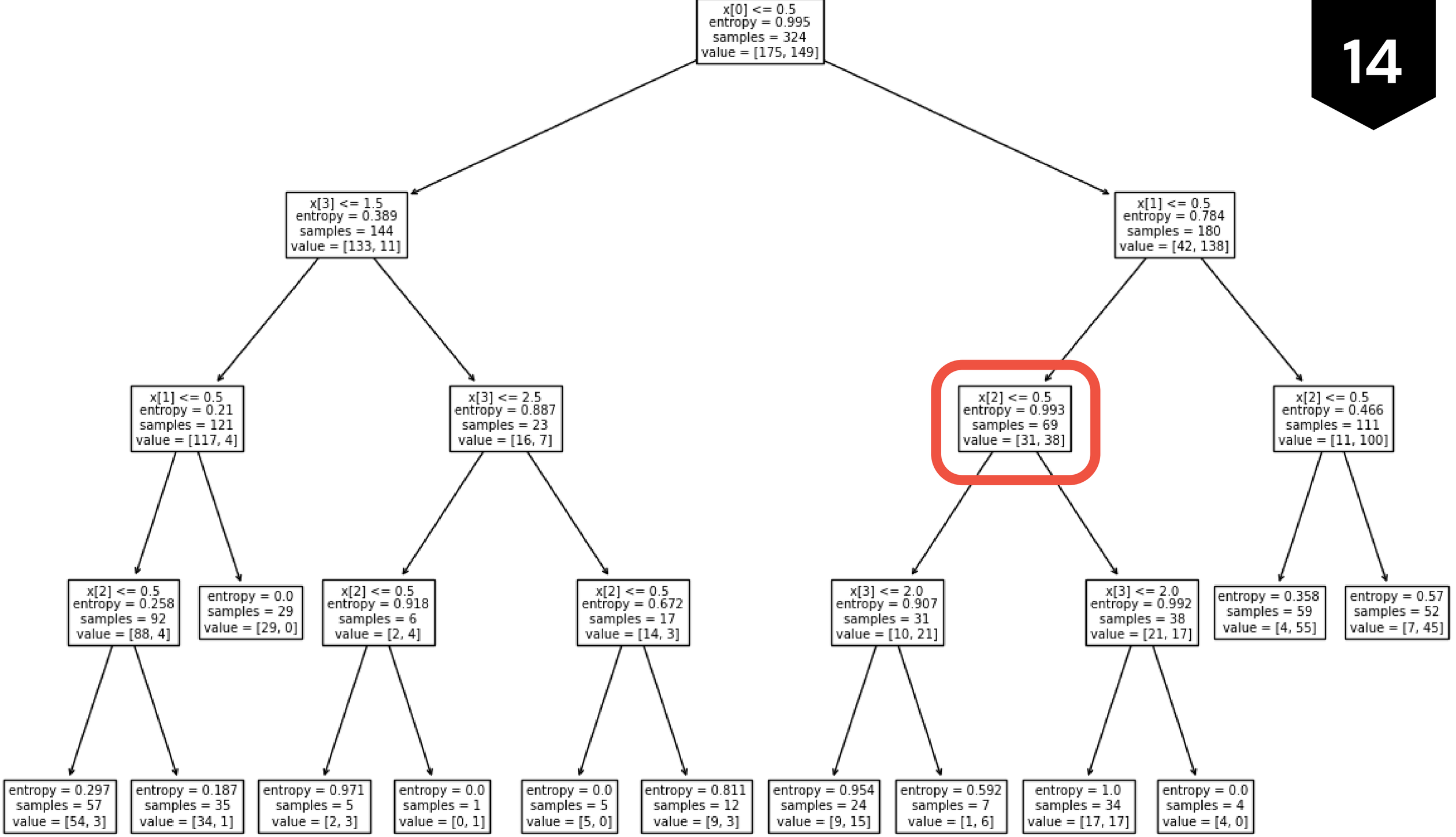
$$\text{Info}(D) = I(31,38)$$

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

$$= 0.992563$$

$$\text{Gain}(a_{12}) = 0.992563 - 0.953881 = \boxed{0.038682}$$

$$\text{Gain}(a_{13}) = 0.992563 - 0.992549 = 0.000014$$


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

$$\text{Info}_{a_{12}}(D) = \frac{31}{69} I(10,21) + \frac{38}{69} I(21,17)$$

$$= \frac{31}{69}\left[-\frac{10}{31}\log_2\frac{10}{31} + \left(-\frac{21}{31}\log_2\frac{21}{31}\right)\right] + \frac{38}{69}\left[-\frac{21}{38}\log_2\frac{21}{38} + \left(-\frac{17}{38}\log_2\frac{17}{38}\right)\right]$$

$$= 0.407567 + 0.546314$$

$$= 0.953881 //$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
0	10	21	31
1	21	17	38
Grand Total	31	38	69

$$\text{Info}_{a_{13}}(D) = \frac{58}{69} I(26,32) + \frac{11}{69} I(5,6)$$

$$= \frac{58}{69}\left[-\frac{26}{58}\log_2\frac{26}{58} + \left(-\frac{32}{58}\log_2\frac{32}{58}\right)\right] + \frac{11}{69}\left[-\frac{5}{11}\log_2\frac{5}{11} + \left(-\frac{6}{11}\log_2\frac{6}{11}\right)\right]$$

$$= 0.83408 + 0.158469$$

$$= 0.992549 //$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
1	26	32	58
3	5	6	11
Grand Total	31	38	69

Splitter “Best”

กรณี $a_9 = 1$ แล: $a_{10} = 1$
หรือ $a_9 \geq 0.5$ แล: $a_{10} \geq 0.5$

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

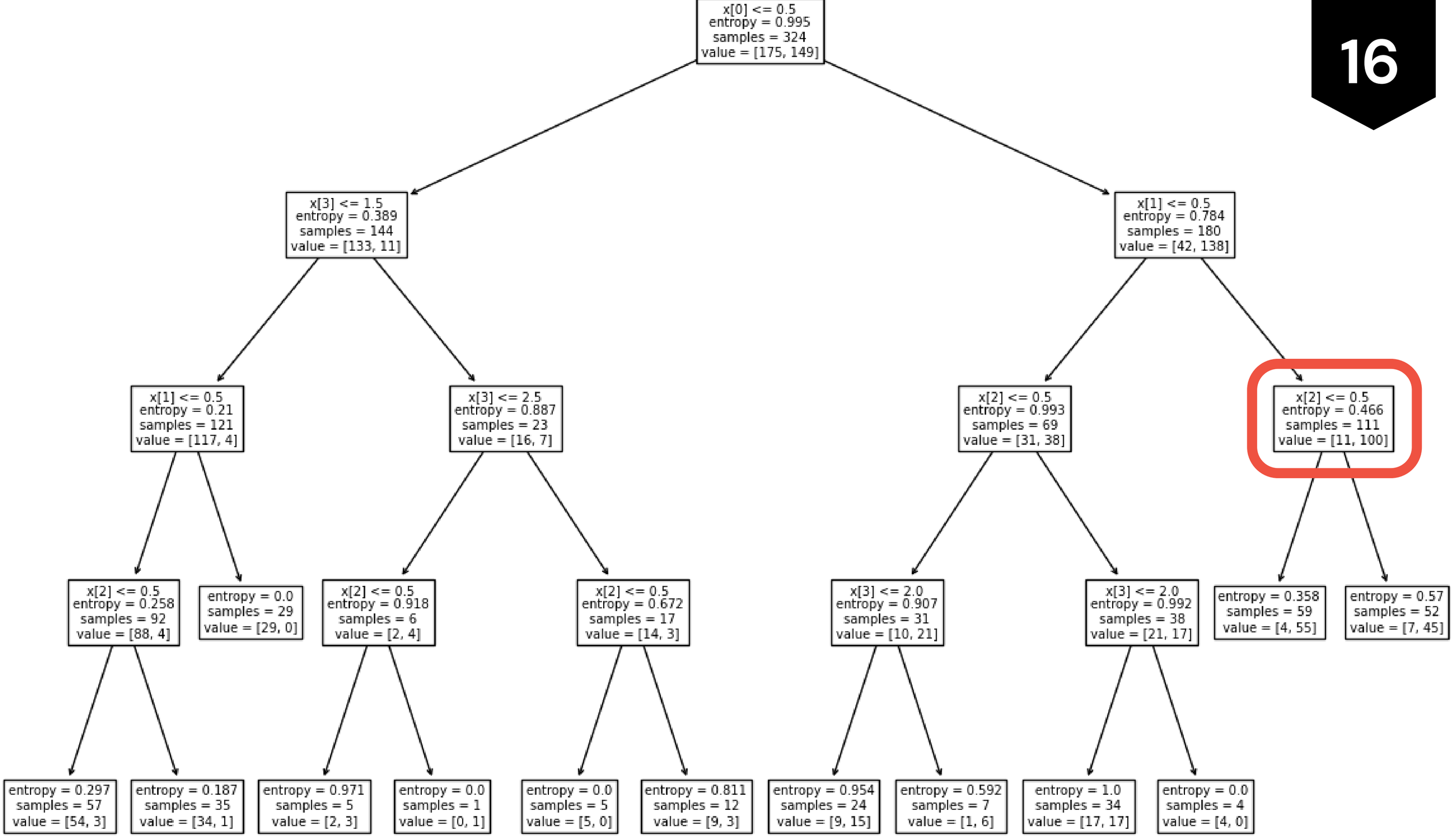
$$= \left[-\frac{11}{111} \log_2 \frac{11}{111} \right] + \left[-\frac{100}{111} \log_2 \frac{100}{111} \right]$$

$$= 0.330494 + 0.135639$$

$$= 0.466133 //$$

$$\text{Gain}(a_{12}) = 0.466133 - 0.457109 = \boxed{0.009024}$$

$$\text{Gain}(a_{13}) = 0.466133 - 0.466133 = 0$$


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

$$\text{Info}_{a_{12}}(D) = \frac{59}{111} I(4,5) + \frac{52}{111} I(45,7)$$

$$= \frac{59}{111}\left[-\frac{4}{59}\log_2\frac{4}{59} + \left(-\frac{5}{59}\log_2\frac{5}{59}\right)\right] + \frac{52}{111}\left[-\frac{45}{52}\log_2\frac{45}{52} + \left(-\frac{7}{52}\log_2\frac{7}{52}\right)\right]$$

$$= 0.1901 + 0.267009$$

$$= 0.457109 //$$

Count of a	Column Labels		
Row La	0	1	Grand Total
0	4	55	59
1	7	45	52
Grand Tot	11	100	111

$$\text{Info}_{a_{13}}(D) = \frac{111}{111} I(11,100)$$

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

$$= 0.330494 + 0.135639$$

$$= 0.466133 //$$

Count of a	Column Labels		
Row La	0	1	Grand Total
1	11	100	111
Grand Tot	11	100	111

Splitter “Best”

กรณี $a_9 = 0$, $a_{13} = 1$, $a_{10} = 0$

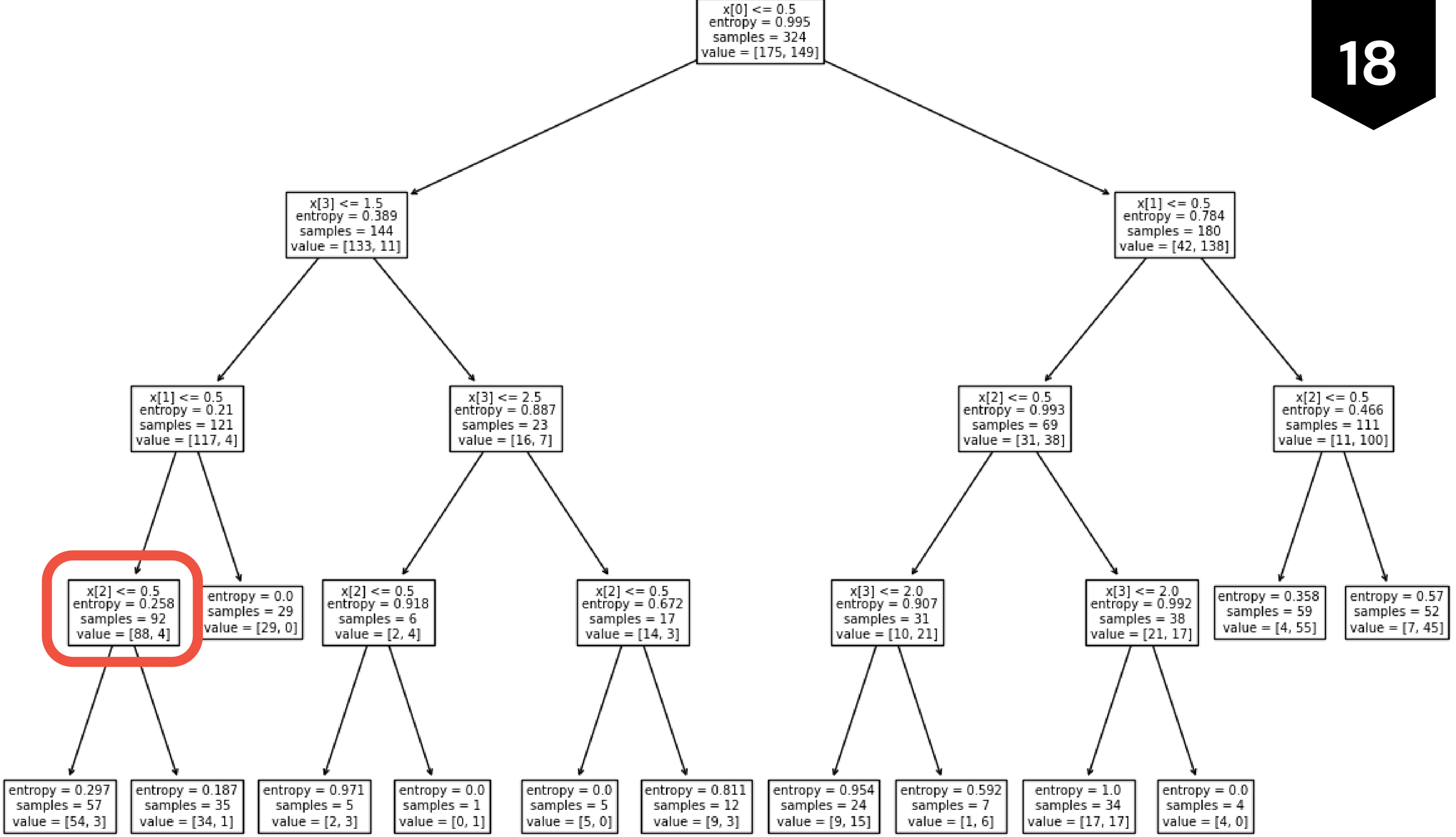
หรือ $a_9 \leq 0.5$, $a_{13} \leq 1.5$, $a_{10} \leq 0.5$

$$Info(D) = I(88,4)$$

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

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

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


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

$$= 0.258019 \text{ //}$$

สิ้นสุดแล้วเนื่องจากใช้ feature, column, dimention, attribute หมดแล้ว

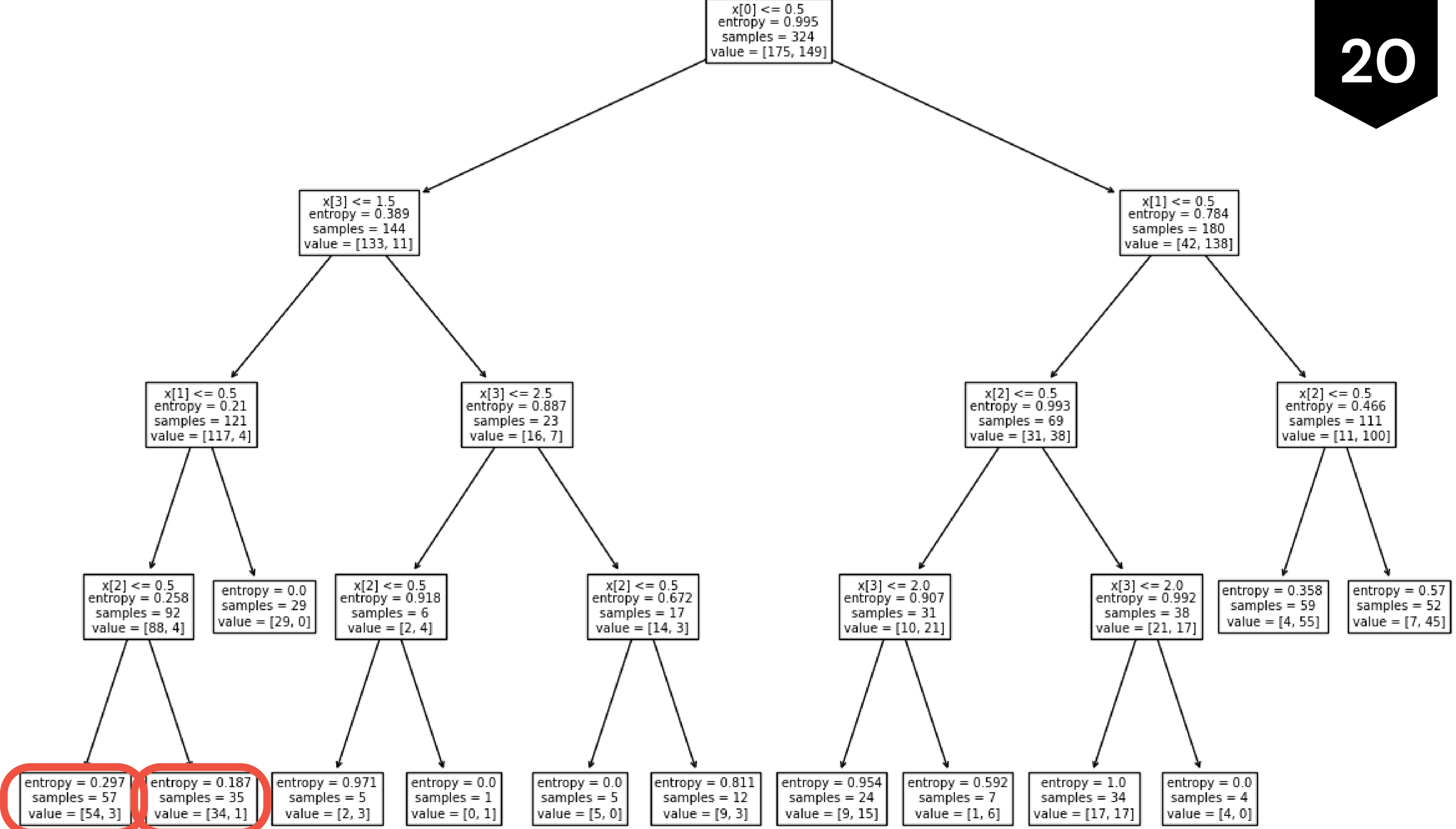
$$Info_{a_{12}}(D) = \frac{57}{92} I(54,3) + \frac{35}{92} I(34,1)$$

$$= \frac{57}{92}\left[-\frac{54}{57}\log_2\frac{54}{57} + \left(-\frac{3}{57}\log_2\frac{3}{57}\right)\right] + \frac{35}{92}\left[-\frac{34}{35}\log_2\frac{34}{35} + \left(-\frac{1}{35}\log_2\frac{1}{35}\right)\right]$$

$$= 0.184303 + 0.0712083$$

$$= 0.2555113 \text{ //}$$

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


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

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

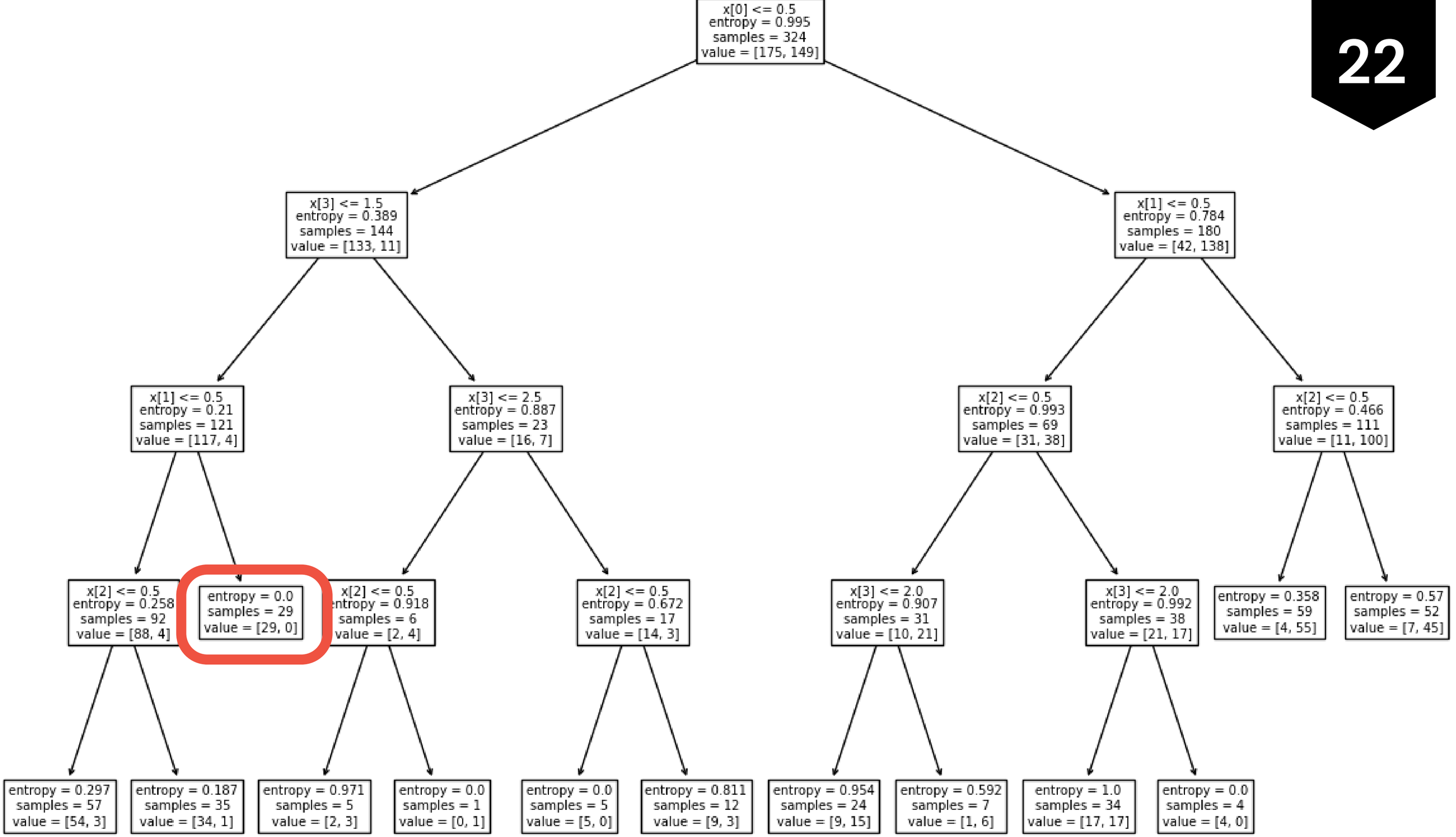
$$Info(D) = I(15,14)$$

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

$$= 0.999142$$

Count of a	Column Labels		
Row La	0	1	Grand Total
0	15	14	29
Grand Tot	15	14	29

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


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

กรณี $a_9 = 0$ และ $a_{13} = (2,3)$ แต่ $a_{13} = 2$

หรือ $a_9 \leq 0.5$ และ $a_{13} \geq 1.5$ แต่ $a_{13} \leq 2.5$

$$Info(D) = I(2,4)$$

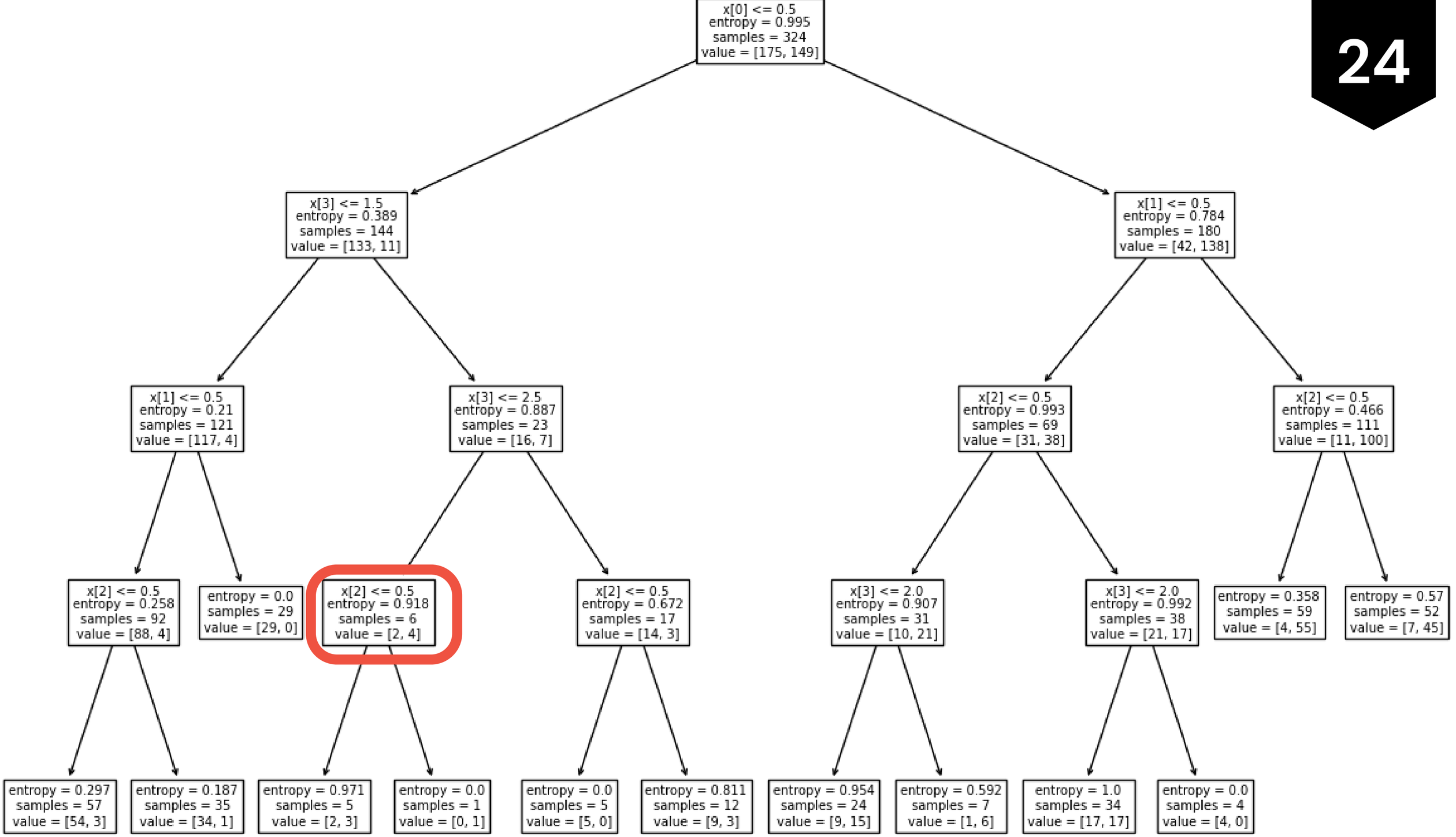
$$= \left[-\frac{2}{6} \log_2 \frac{2}{6} \right] + \left[-\frac{4}{6} \log_2 \frac{4}{6} \right]$$

$$= 0.918296 \text{ //}$$

↗ 0

Column Labels	0	1	Grand Total
	2	3	5
		1	1
	2	**4**	**6**

$$Gain(a_{12}) = 0.918296 - 0.809126 = \boxed{0.10917}$$


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

$$= 0.918296$$

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

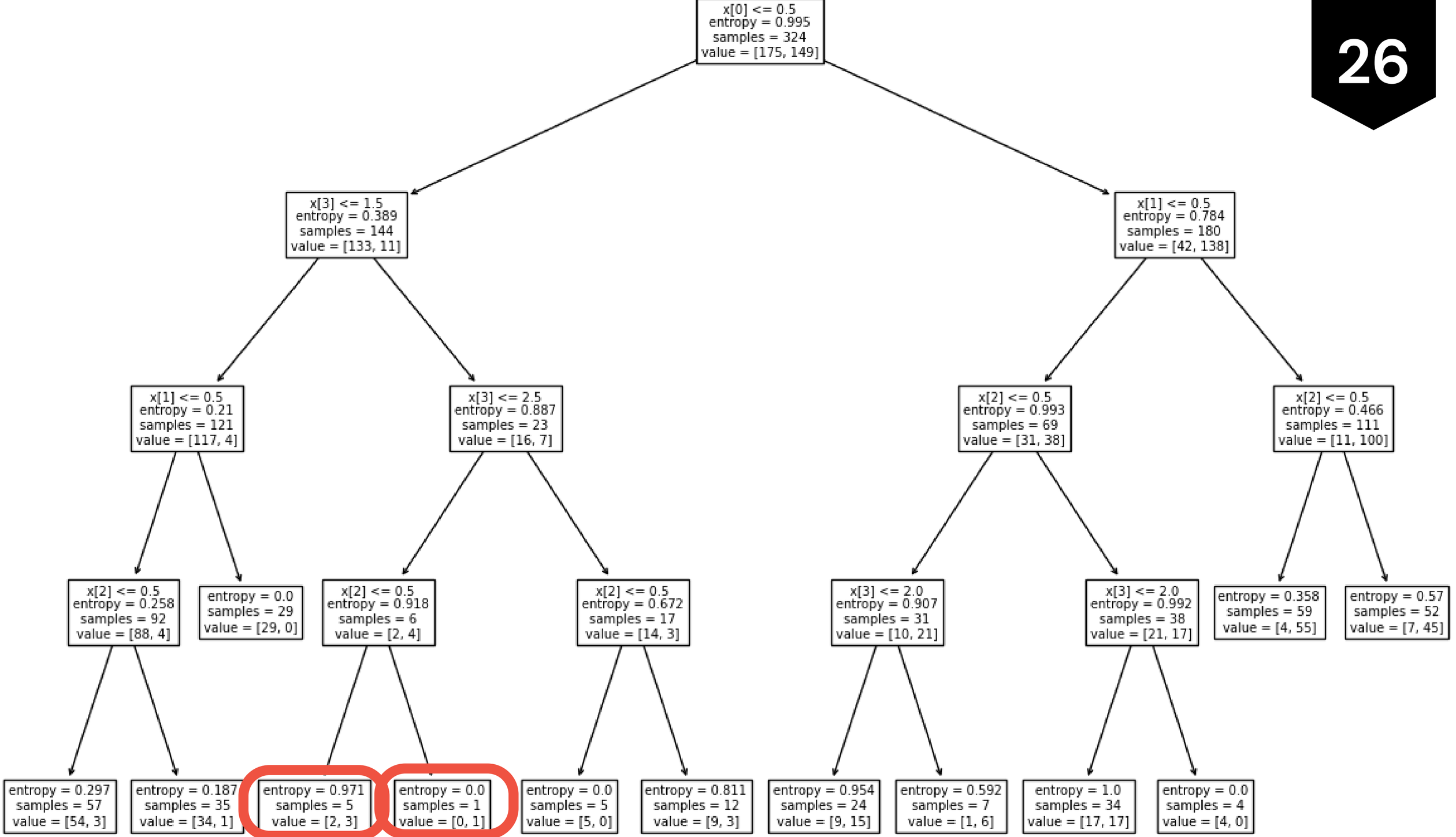
0.97095

$$= \frac{5}{6}\left[-\frac{2}{5}\log_2\frac{2}{5} + \left(-\frac{3}{5}\log_2\frac{3}{5}\right)\right] + \frac{1}{6}\left[-\frac{0}{1}\log_2\frac{0}{1} + \left(-\frac{1}{1}\log_2\frac{1}{1}\right)\right]$$

หาทำไม่ได้ 0

$$= 0.809126$$

Count of a Column Labels			
Row La	0	1	Grand Total
0	5		5
1	9	3	12
Grand Tot	14	3	17


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

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

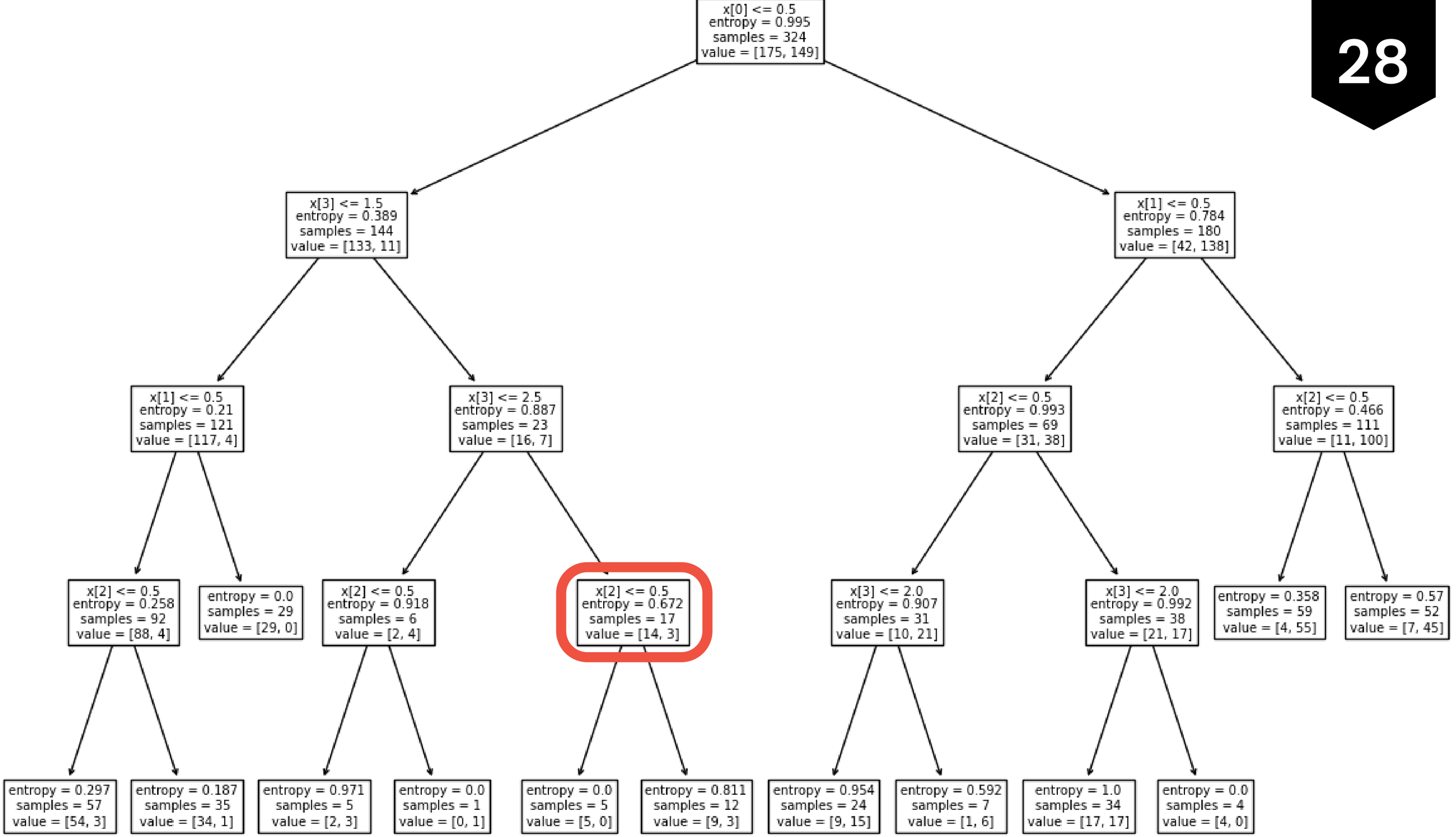
$$\text{Info}(D) = I(14,3)$$

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

$$= 0.672295 \text{ //}$$

สิ้นสุดแล้ว

$$\text{Gain}(a_{12}) = 0.672295 - 0.57267 = \boxed{0.099628}$$


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

$= 0.672295$ //

สิ้นสุดแล้ว

$$\text{Info}_{a_{12}}(D) = \frac{5}{17} I(5,0) + \frac{12}{17} I(9,3)$$

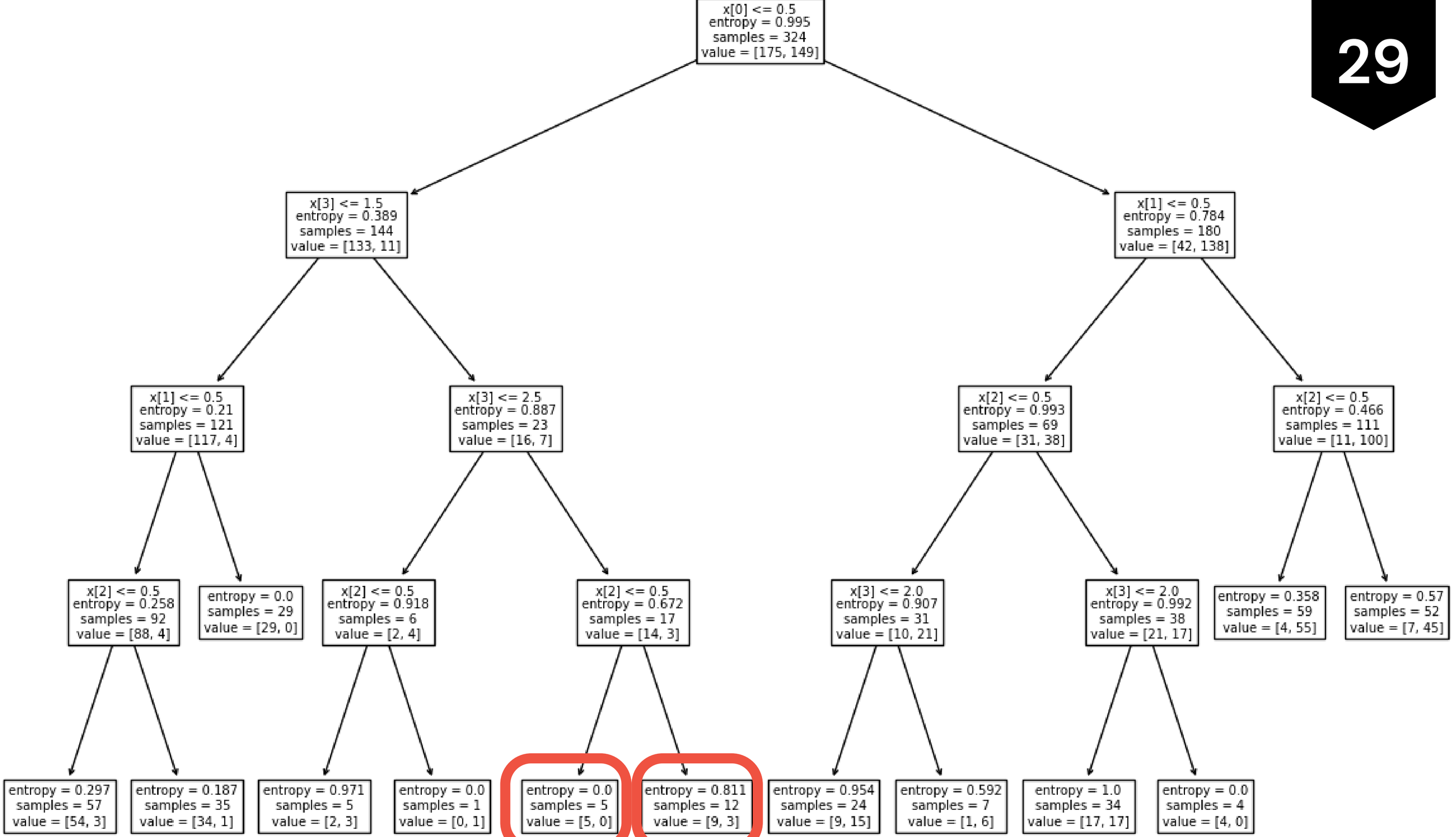
0 หาค่าไม่ได้

$$= \frac{5}{17}\left[-\frac{5}{5}\log_2\frac{5}{5} + \left(-\frac{0}{5}\log_2\frac{0}{5}\right)\right] + \frac{12}{17}\left[-\frac{9}{12}\log_2\frac{9}{12} + \left(-\frac{3}{12}\log_2\frac{3}{12}\right)\right]$$

$= 0.572667$ //

0.811278

$$\text{Gain}(a_{12}) = 0.672295 - 0.57667 = 0.099628$$


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

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

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

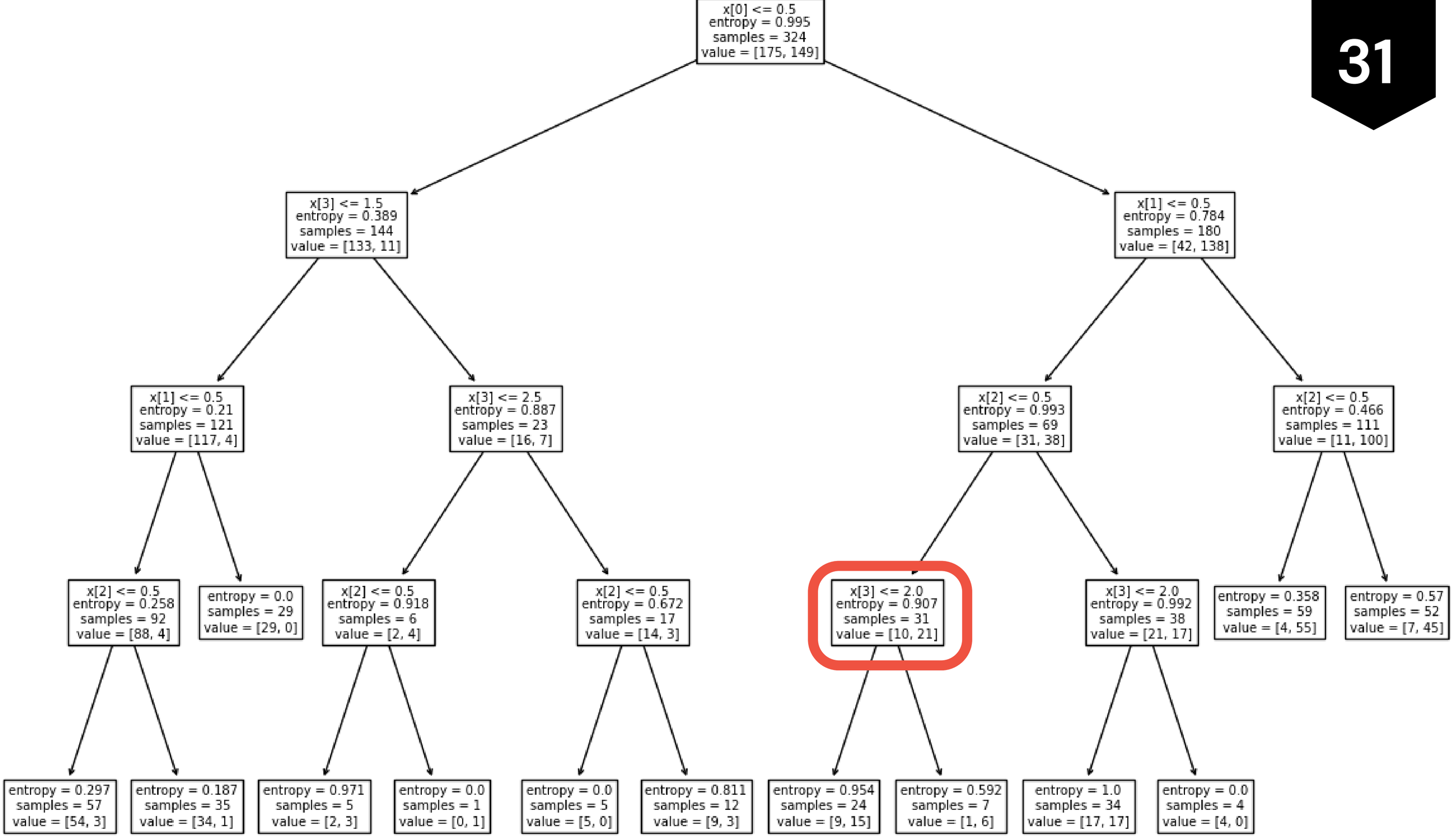
$$= \left[-\frac{10}{31} \log_2 \frac{10}{31} \right] + \left[-\frac{21}{31} \log_2 \frac{21}{31} \right]$$

$$= 0.526538 + 0.380628$$

$$= 0.907166$$

Column Labels	0	1	Grand Total
	4	55	59
	4	55	59

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


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter "Best"

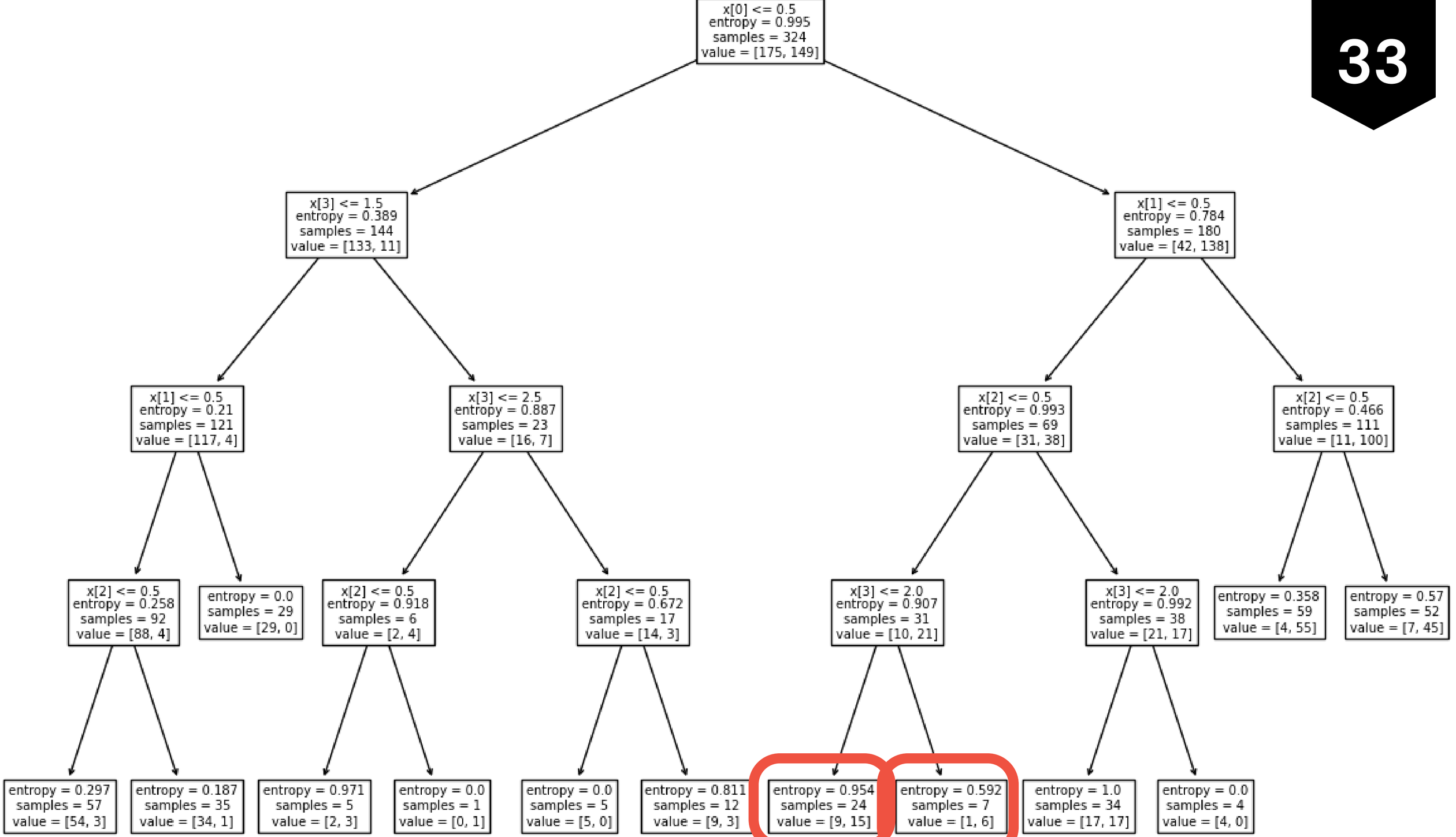
$$\text{Info}_{a_{13}}(D) = \frac{24}{31} I(9, 15) + \frac{7}{31} I(1, 6)$$

$$= \frac{24}{31}\left[-\frac{9}{24}\log_2\frac{9}{24} + \left(-\frac{15}{24}\log_2\frac{15}{24}\right)\right] + \frac{7}{31}\left[-\frac{1}{7}\log_2\frac{1}{7} + \left(-\frac{6}{7}\log_2\frac{6}{7}\right)\right]$$

$$= 0.738917 + 0.133604$$

$$= 0.872521 \text{ //}$$

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


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

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

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

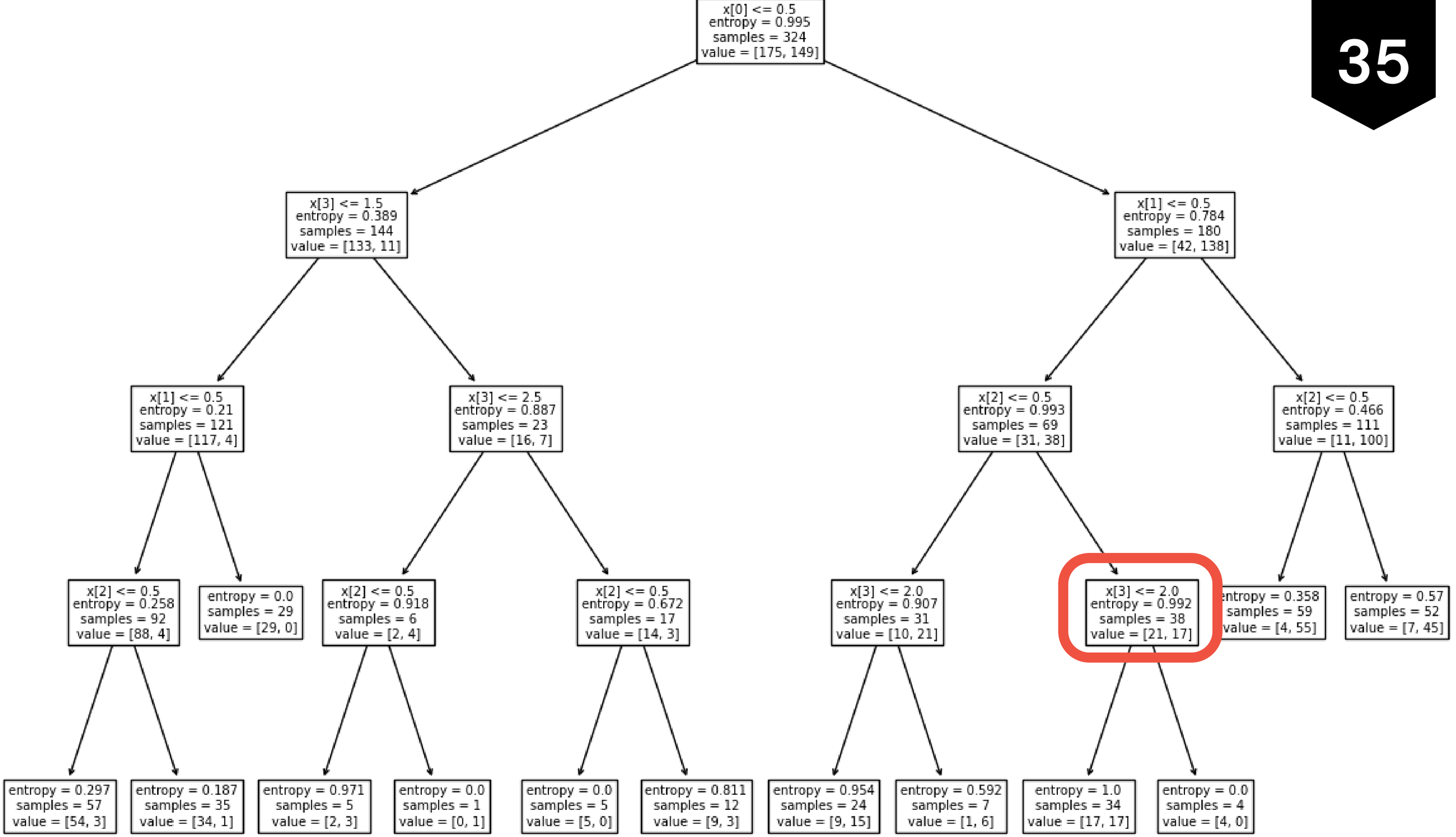
$$= \left[-\frac{21}{38} \log_2 \frac{21}{38} \right] + \left[-\frac{17}{38} \log_2 \frac{17}{38} \right]$$

$$= 0.472837 + 0.519155$$

$$= 0.991992 \text{ //}$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
0	10	21	31
1	21	17	38
Grand Total	31	38	69

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


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

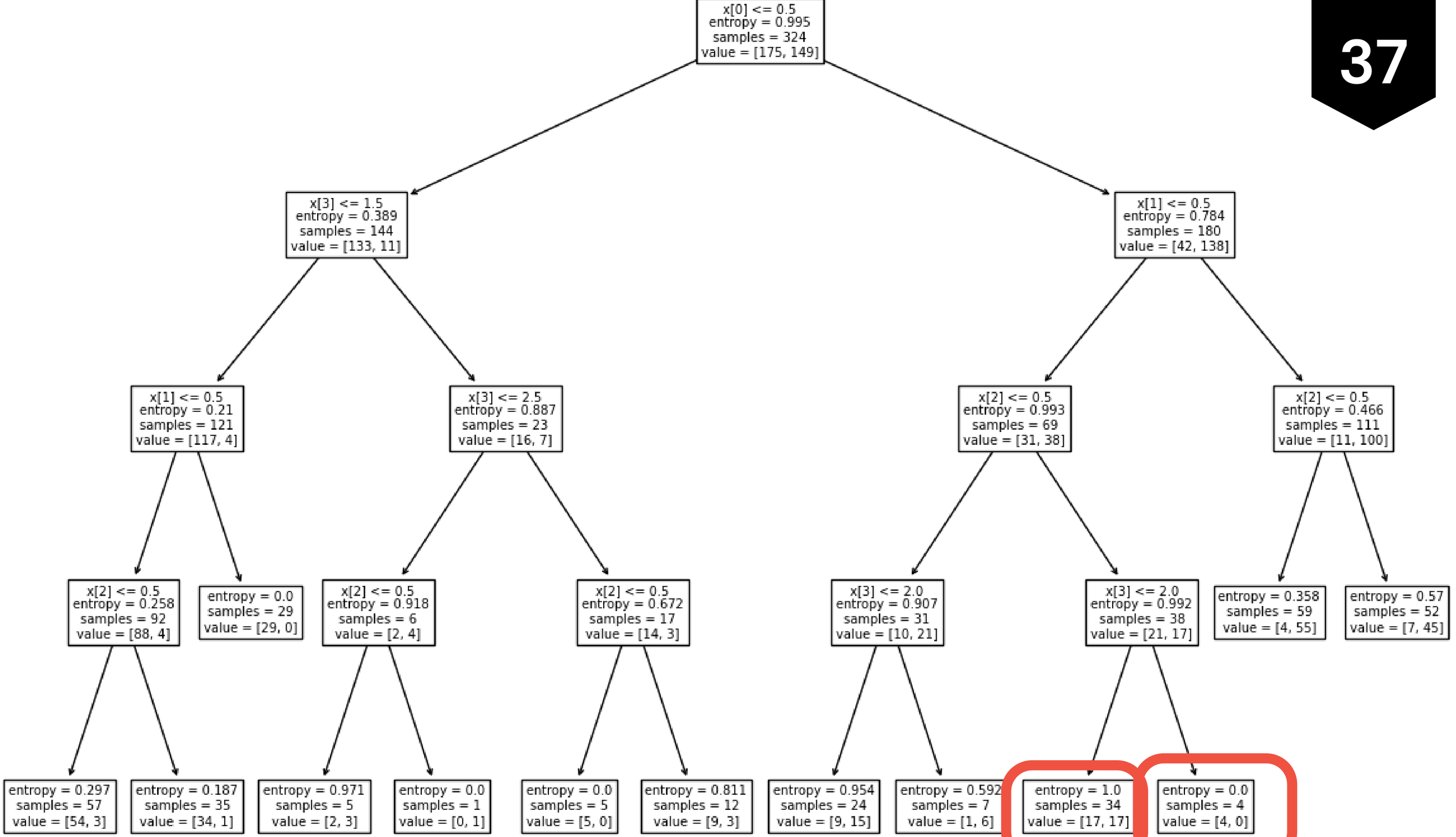
$$\text{Info}_{a_{13}}(D) = \frac{34}{38} I(17,17) + \frac{4}{36} I(4,0)$$

$$= \frac{34}{38}\left[-\frac{17}{34}\log_2\frac{17}{34} + \left(-\frac{17}{34}\log_2\frac{17}{34}\right)\right] + \frac{4}{36}\left[-\frac{4}{4}\log_2\frac{4}{4} + \left(-\frac{0}{4}\log_2\frac{0}{4}\right)\right]$$

0 (ขีดฆ่า) หาไม่ได้ (ขีดฆ่า)

$$= \frac{34}{38} \times (0.5 + 0.5)$$

$$= 0.894737 \text{ //}$$


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

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

$$Info(D) = I(4,55)$$

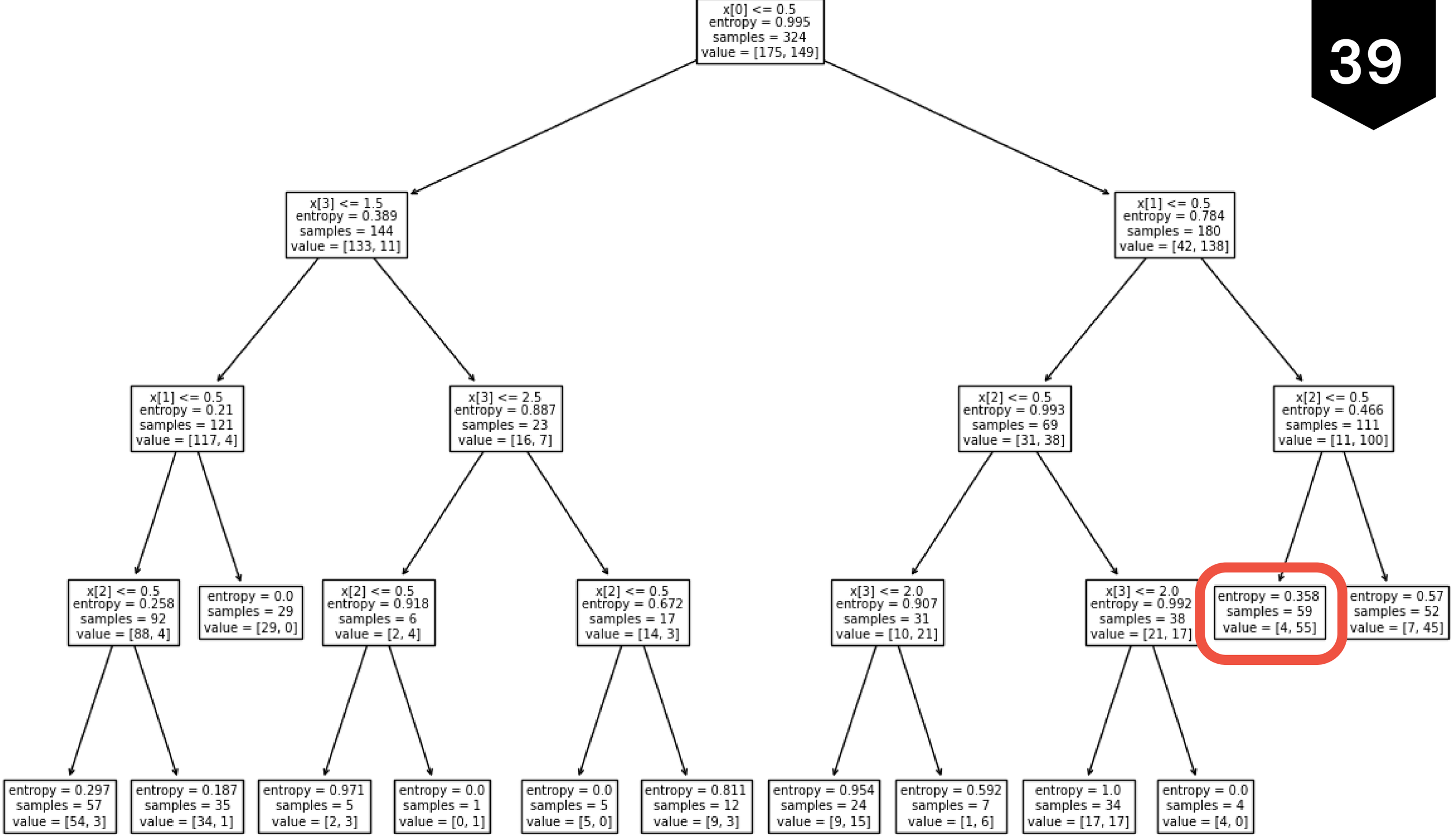
$$= \left[-\frac{4}{59} \log_2 \frac{4}{59} \right] + \left[-\frac{55}{59} \log_2 \frac{55}{59} \right]$$

$$= 0.26323 + 0.0944167$$

$$= 0.3576467$$

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

ไม่มีการแบ่งต่อเพราะค่าที่เป็นไปได้ค่าเดียวคือ 1


x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]

Splitter “Best”

กรณี $a_9 = 1$, $a_{10} = 1$, $a_{12} = 1$
หรือ $a_9 \geq 0.5$, $a_{10} \geq 0.5$, $a_{12} \geq 0.5$

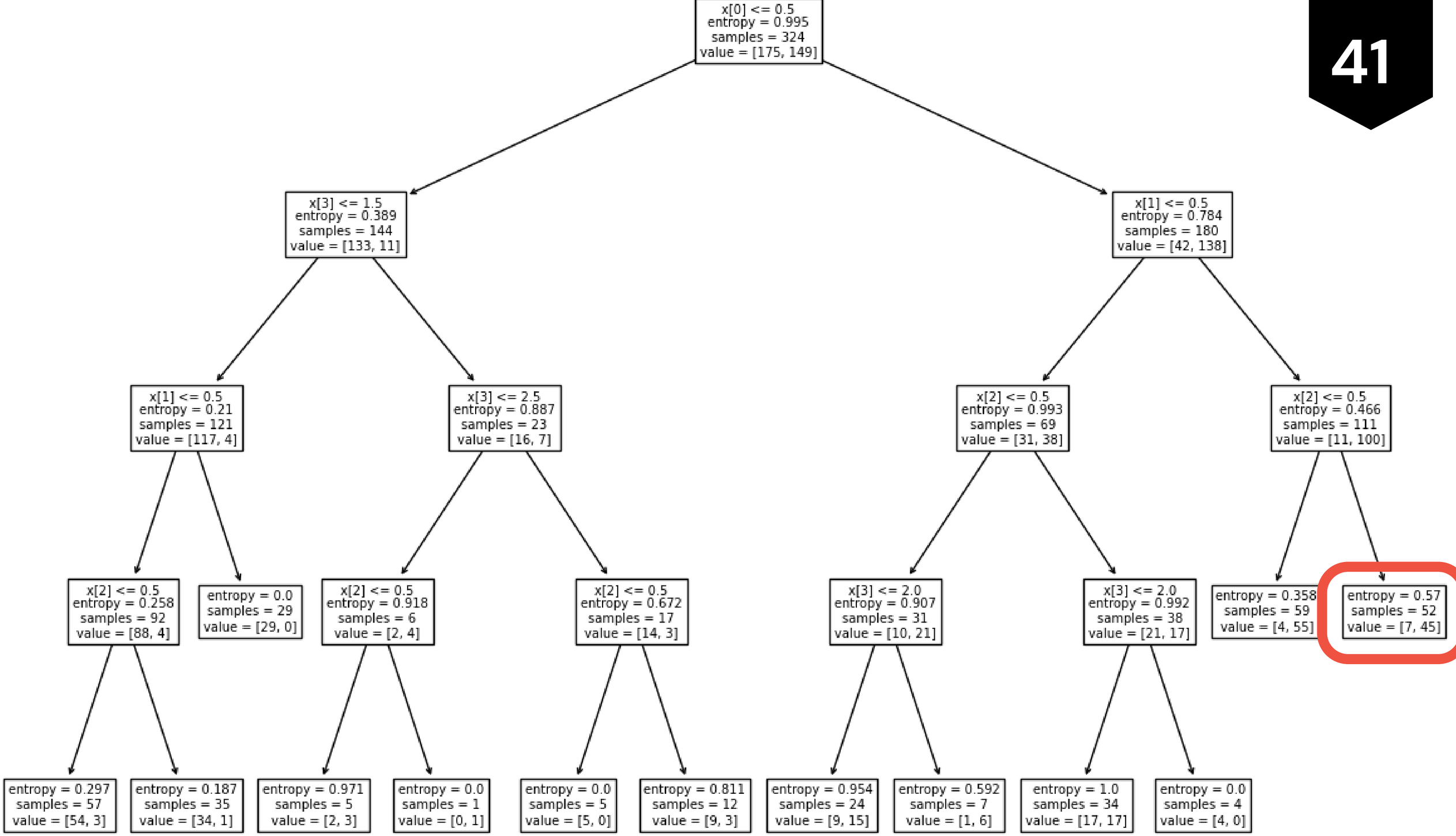
$$Info(D) = I(7,45)$$

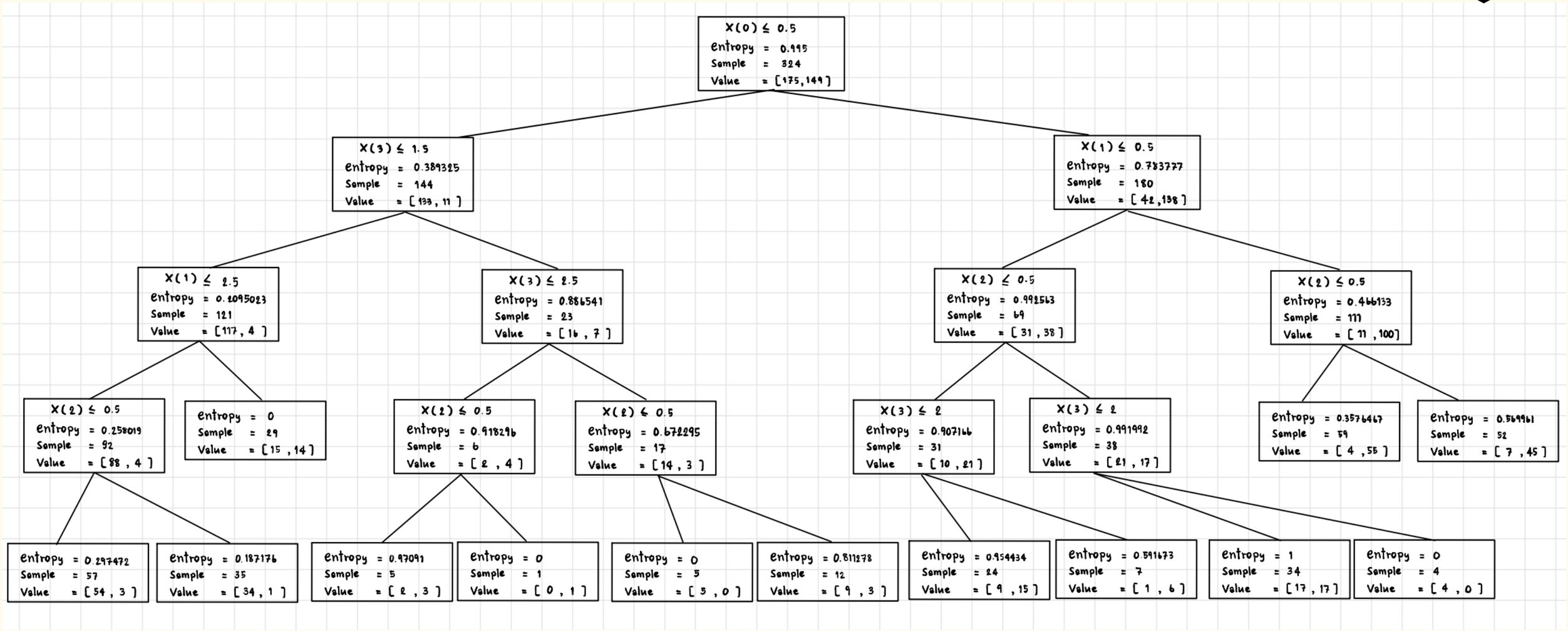
$$= \left[-\frac{7}{52} \log_2 \frac{7}{52} \right] + \left[-\frac{45}{52} \log_2 \frac{45}{52} \right]$$

$$= 0.569961$$

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

ไม่มีการแบ่งต่อเพราะค่าที่เป็นไปได้ค่าเดียวคือ 1

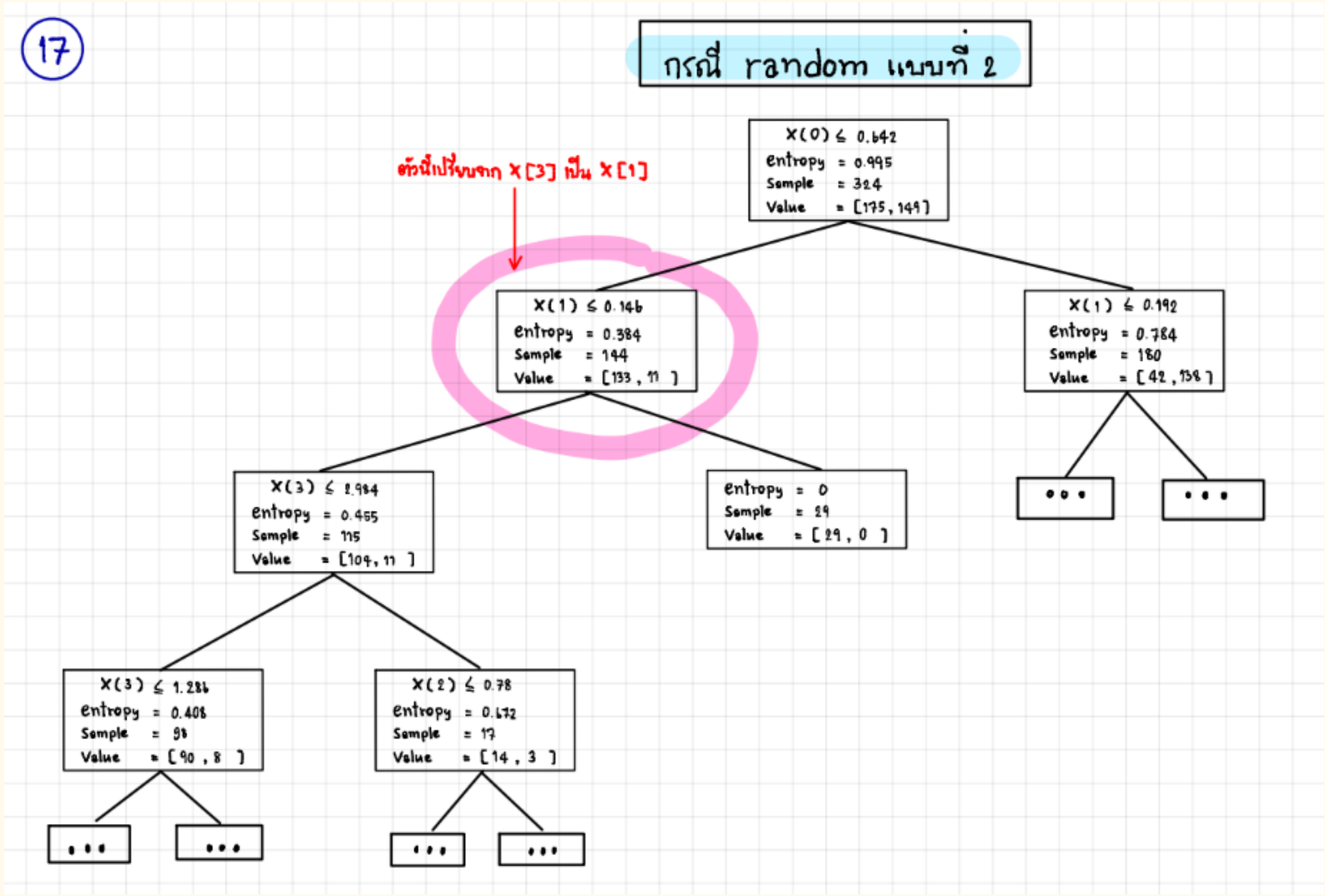

x[0] <= 0.5
entropy = 0.995
samples = 324
value = [175, 149]
x[3] <= 1.5
entropy = 0.389
samples = 144
value = [133, 11]
x[1] <= 0.5
entropy = 0.784
samples = 180
value = [42, 138]
x[1] <= 0.5
entropy = 0.21
samples = 121
value = [117, 4]
x[3] <= 2.5
entropy = 0.887
samples = 23
value = [16, 7]
x[2] <= 0.5
entropy = 0.993
samples = 69
value = [31, 38]
x[2] <= 0.5
entropy = 0.466
samples = 111
value = [11, 100]
x[2] <= 0.5
entropy = 0.258
samples = 92
value = [88, 4]
entropy = 0.0
samples = 29
value = [29, 0]
x[2] <= 0.5
entropy = 0.918
samples = 6
value = [2, 4]
x[2] <= 0.5
entropy = 0.672
samples = 17
value = [14, 3]
x[3] <= 2.0
entropy = 0.907
samples = 31
value = [10, 21]
x[3] <= 2.0
entropy = 0.992
samples = 38
value = [21, 17]
entropy = 0.358
samples = 59
value = [4, 55]
entropy = 0.57
samples = 52
value = [7, 45]
entropy = 0.297
samples = 57
value = [54, 3]
entropy = 0.187
samples = 35
value = [34, 1]
entropy = 0.971
samples = 5
value = [2, 3]
entropy = 0.0
samples = 1
value = [0, 1]
entropy = 0.0
samples = 5
value = [5, 0]
entropy = 0.811
samples = 12
value = [9, 3]
entropy = 0.954
samples = 24
value = [9, 15]
entropy = 0.592
samples = 7
value = [1, 6]
entropy = 1.0
samples = 34
value = [17, 17]
entropy = 0.0
samples = 4
value = [4, 0]



ความแม่นยำ = 83.88 %

Splitter “Random”

Splitter “Random”



Splitter “Random”

random ต้นไม้ที่เป็นไปได้

กรณี $a_9 \le 0.299$ แล้ว $a_{10} \le 0.744$
หรือ $a_9 = 0$ แล้ว $a_{10} = 0$

$$\text{Info}(D) = I(104, 11)$$

$$= \left[-\frac{104}{115} \log_2 \frac{104}{115} \right] + \left[-\frac{11}{115} \log_2 \frac{11}{115} \right]$$

$$= 0.131176 + 0.323884$$

$$= 0.45506 //$$

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

$$\text{Gain}(a_{13}) \rightarrow = 0.45506 - 0.383726 = 0.0678 \rightarrow \text{กรณี } (1,(2,3))$$

$$\text{Gain}(a_{13}) \rightarrow = 0.45506 - 0.4469897 = 0.0080703 \rightarrow \text{กรณี } ((1,2),3)$$

Splitter “Random”

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

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

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

$$\text{Info}_{a_{13}}(D) = \frac{92}{115} I(88,4) + \frac{23}{115} I(16,7)$$

กรณี (1,(2,3))

$$= \frac{92}{115}\left[-\frac{88}{92}\log_2\frac{88}{92} + \left(-\frac{4}{92}\log_2\frac{4}{92}\right)\right] + \frac{23}{115}\left[-\frac{16}{23}\log_2\frac{16}{23} + \left(-\frac{7}{23}\log_2\frac{7}{23}\right)\right]$$
$$= 0.206415 + 0.177311$$
$$= 0.383726 //$$

$$\text{Info}_{a_{13}}(D) = \frac{98}{115} I(90,8) + \frac{17}{115} I(14,3)$$

กรณี ((1,2),3)

$$= \frac{98}{115}\left[-\frac{90}{98}\log_2\frac{90}{98} + \left(-\frac{8}{98}\log_2\frac{8}{98}\right)\right] + \frac{17}{115}\left[-\frac{14}{17}\log_2\frac{14}{17} + \left(-\frac{3}{17}\log_2\frac{3}{17}\right)\right]$$
$$= 0.347607 + 0.0993827$$
$$= 0.4469897$$

Splitter “Random”

19

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

$$\text{Info}(D) = I(90, 8)$$

$$= \left[-\frac{90}{98} \log_2 \frac{90}{98} \right] + \left[-\frac{8}{98} \log_2 \frac{8}{98} \right]$$

$$= 0.112828 + 0.295078$$

$$= 0.407906 \text{ //}$$

Count of a16	Column Labels		
Row Labels	0	1	Grand Total
0	56	6	62
1	34	2	36
Grand Total	**90**	**8**	**98**

$$\text{Info}_{a_{12}}(D) = \frac{62}{98} I(56, 6) + \frac{36}{98} I(34, 2)$$

$$= \frac{62}{98}\left[-\frac{56}{62} \log_2 \frac{56}{62} + \left(-\frac{6}{62} \log_2 \frac{6}{62} \right) \right] + \frac{36}{98}\left[-\frac{34}{36} \log_2 \frac{34}{36} + \left(-\frac{2}{36} \log_2 \frac{2}{36} \right) \right]$$

$$= 0.290189 + 0.11371$$

$$= 0.403899$$

Splitter “Random”

$$\text{Info}\ a_{13}(D) = \frac{92}{98} I(88,4) + \frac{6}{98} I(2,4)$$

กรณี (1,2)

$$= \frac{92}{98}\left[-\frac{88}{92}\log_2\frac{88}{92} + \left(-\frac{4}{92}\log_2\frac{4}{92}\right)\right] + \frac{6}{98}\left[-\frac{2}{6}\log_2\frac{2}{6} + \left(-\frac{4}{6}\log_2\frac{4}{6}\right)\right]$$

$$= 0.242222 + 0.056222$$

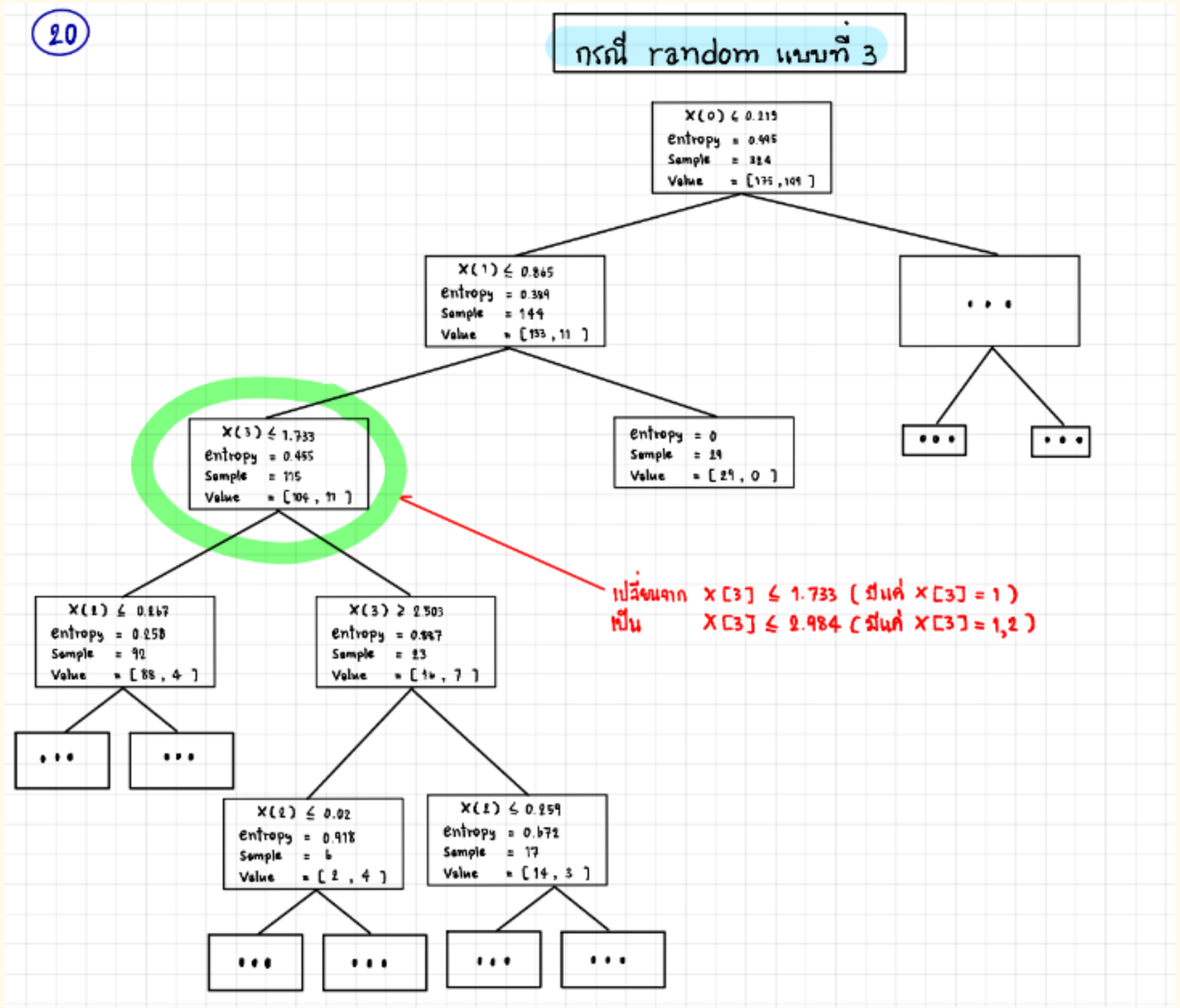
$$= 0.298442$$

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

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

$$= \left[-\frac{16}{23} \log_2 \frac{16}{23} \right] + \left[-\frac{7}{23} \log_2 \frac{7}{23} \right]$$

$$= 0.364217 + 0.522324$$

$$= 0.886541$$

$$\text{Gain}(a_{12}) = 0.886541 - 0.886492 = 0.000049$$

$$\text{Gain}(a_{13}) = 0.886541 - 0.736469 = \boxed{0.150072}$$

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**

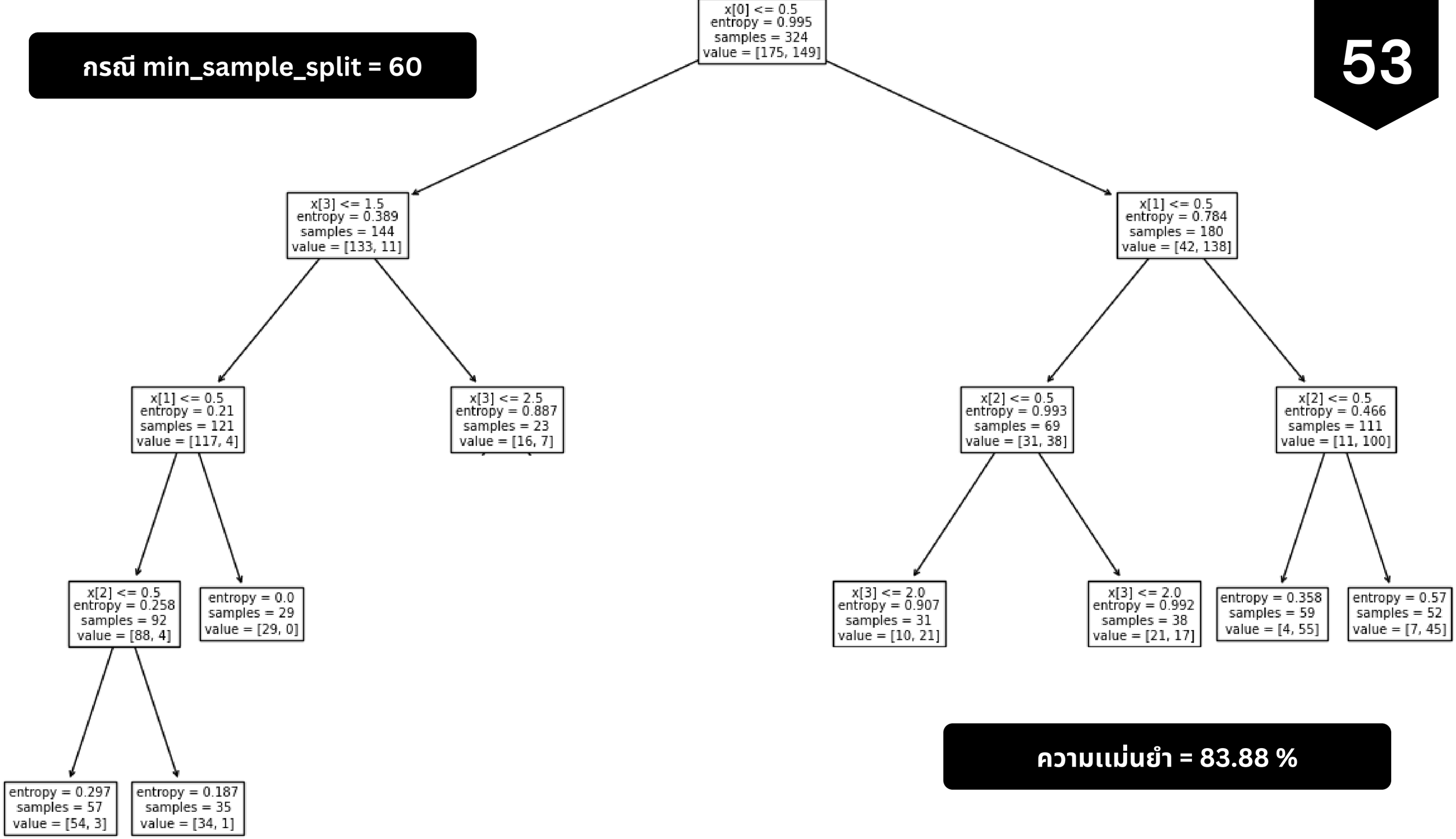
Splitter “Random”

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

Count of a	Column Labels		
Row La	0	1	Grand Total
2	2	4	6
3	14	3	17
Grand Tot	16	7	23

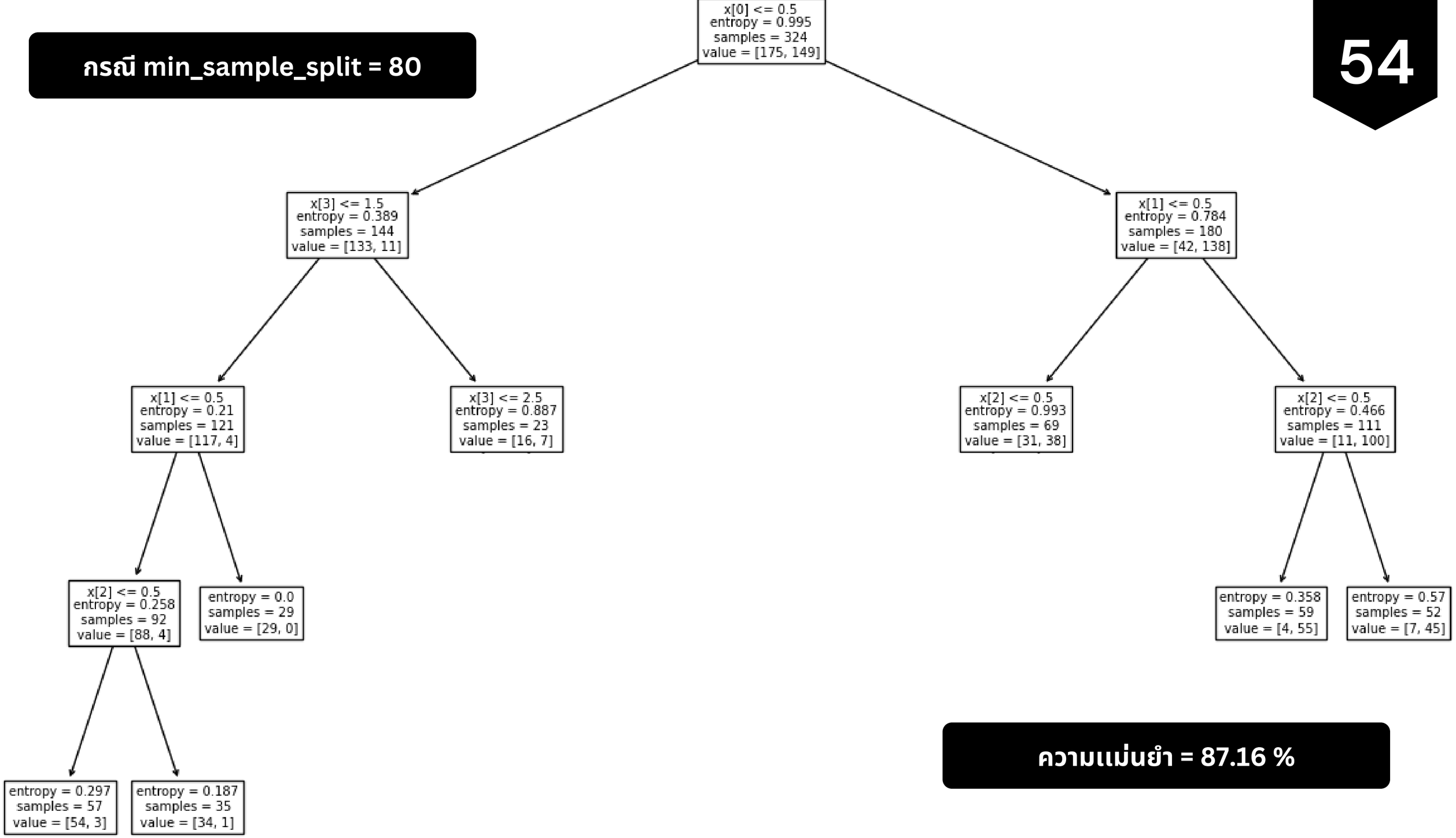
Min Samples Split

กรณี min_sample_split = 60



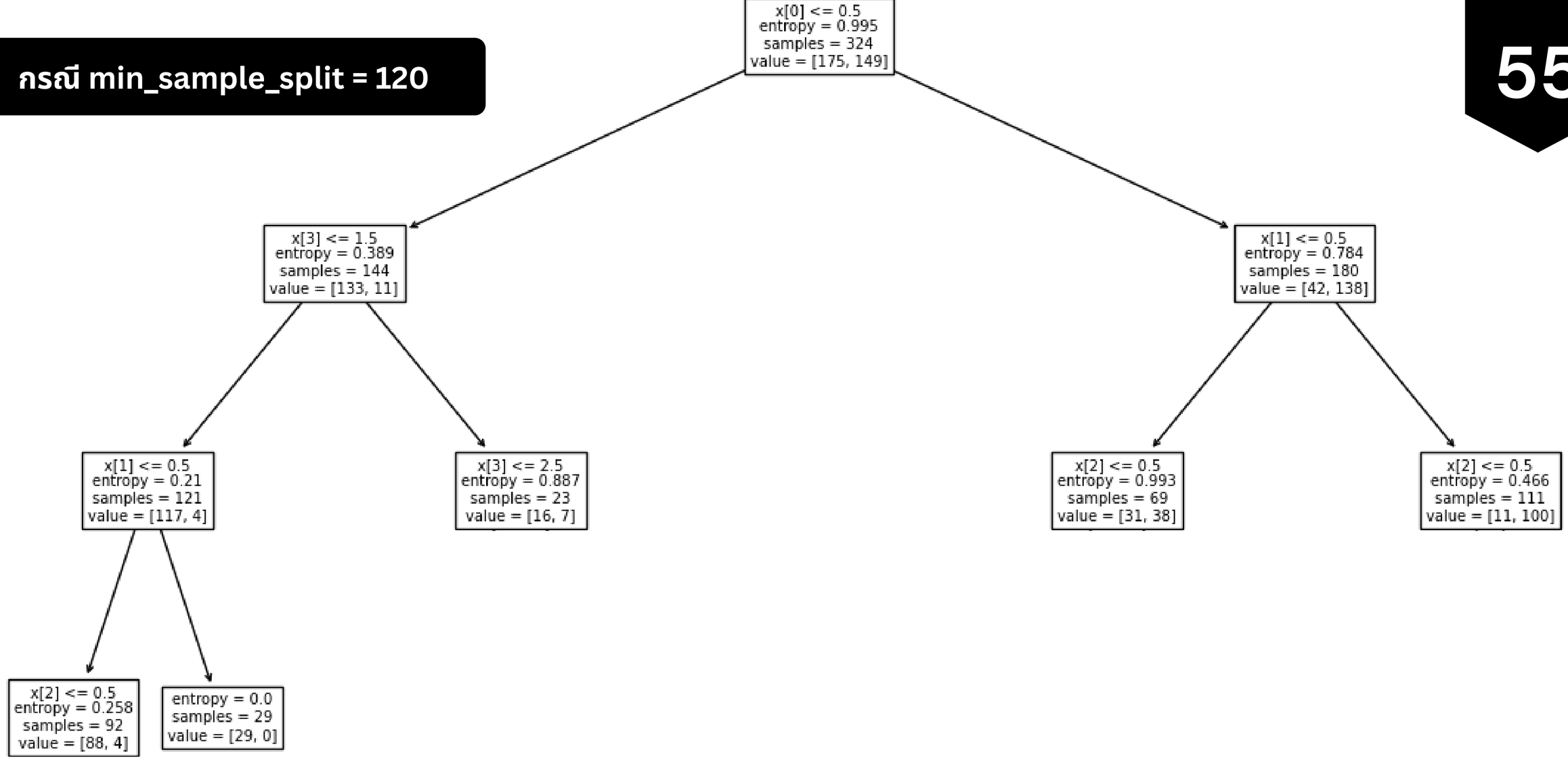
ความแม่นยำ = 83.88 %

กรณี min_sample_split = 80



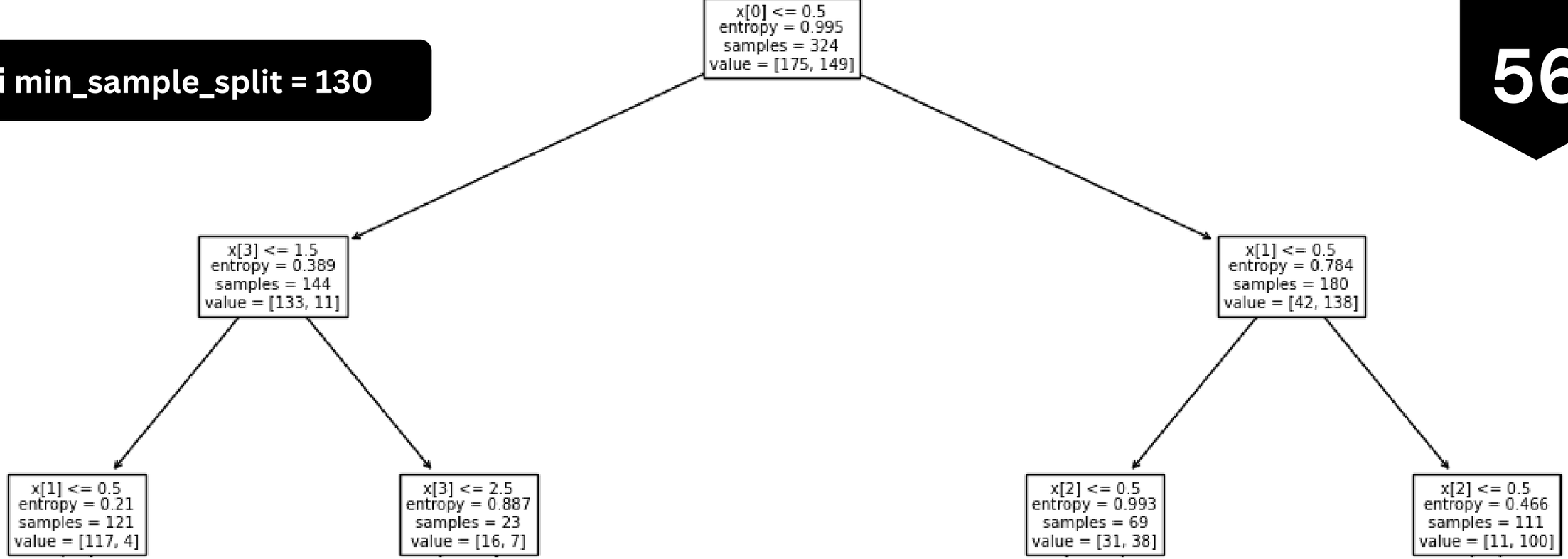
ความแม่นยำ = 87.16 %

กรณี min_sample_split = 120



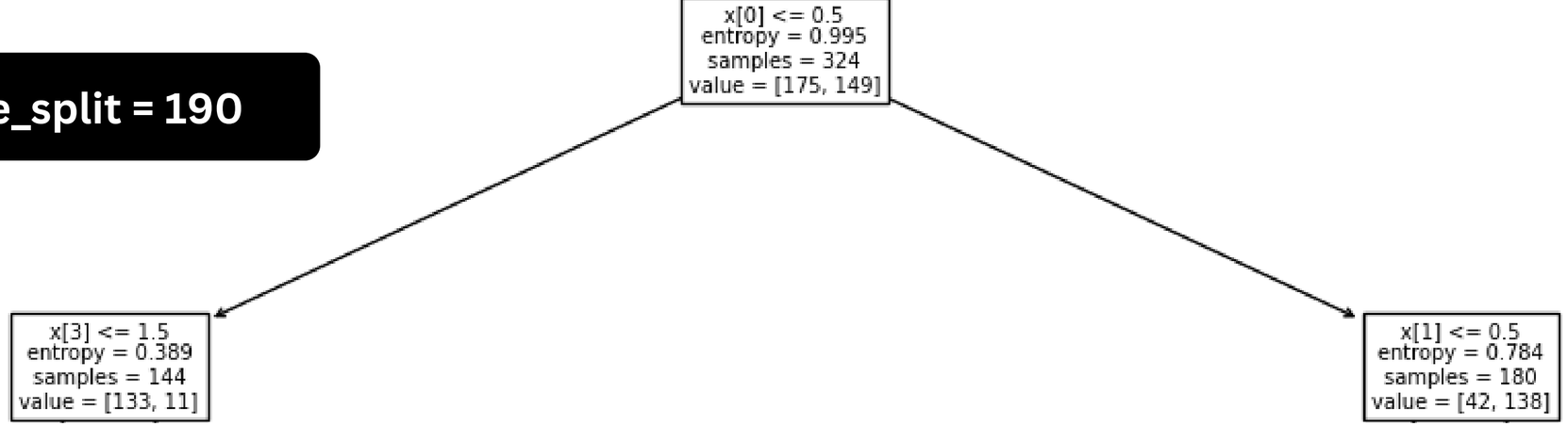
ความแม่นยำ = 87.16 %

กรณี min_sample_split = 130



ความแม่นยำ = 87.16 %

กรณี min_sample_split = 190



ความแม่นยำ = 87.16 %


THANK YOU