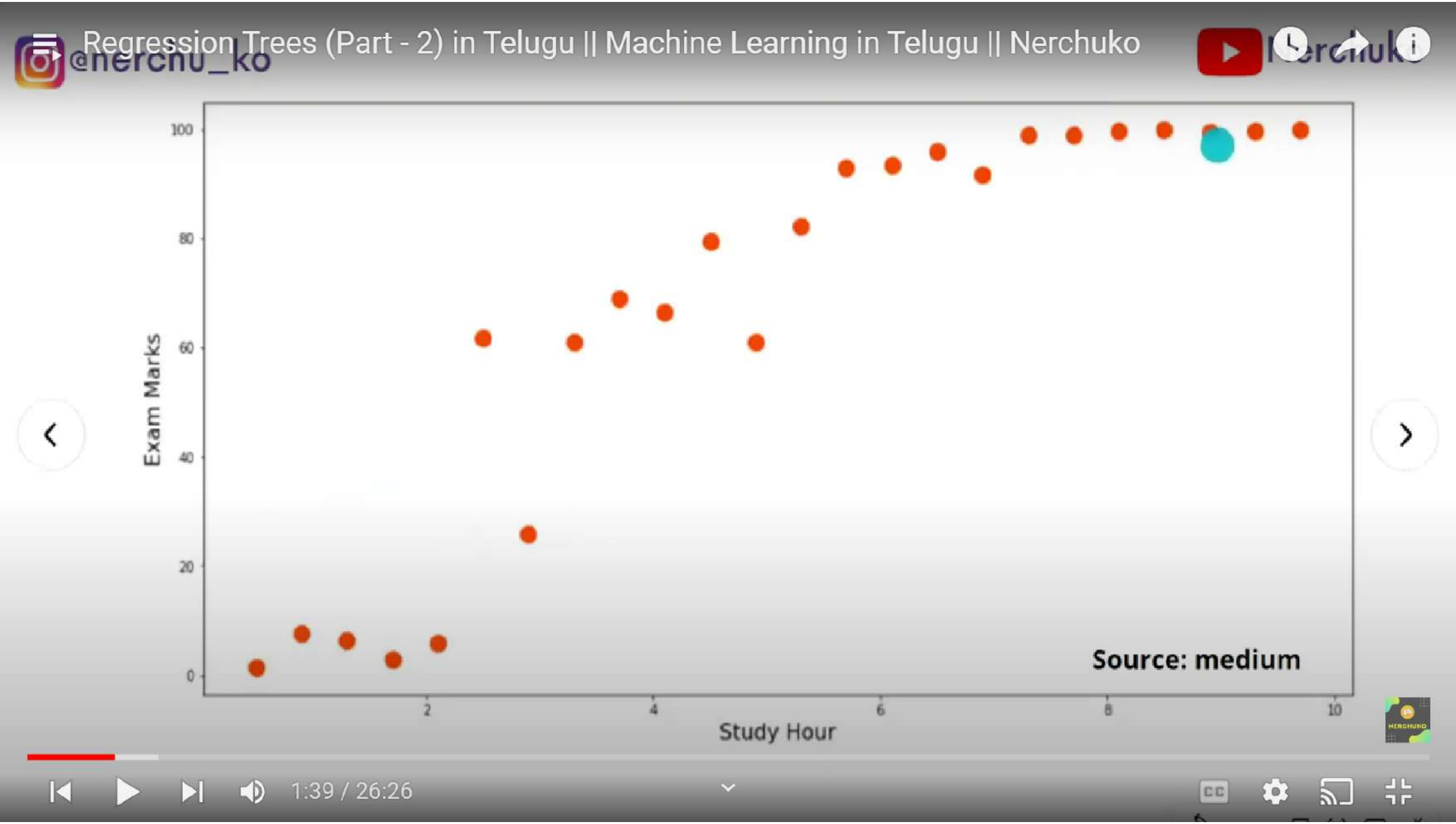
@nerchu_ko

Nerchuko

	Study Hours	Marks
0	0.5	1.4
1	0.9	7.6
2	1.3	6.4
3	1.7	2.8
4	2.1	5.9
5	2.5	61.7
6	2.9	25.9
7	3.3	61.1
8	3.7	69.1
9	4.1	66.6
10	4.5	79.5
11	4.9	61.0
12	5.3	82.3
13	5.7	92.9
14	6.1	93.4
15	6.5	96.0
16	6.9	91.7
17	7.3	98.9
18	7.7	98.8
19	8.1	99.7
20	8.5	100.0
21	8.9	99.5
22	9.3	99.6
23	9.7	99.9

Source: medium

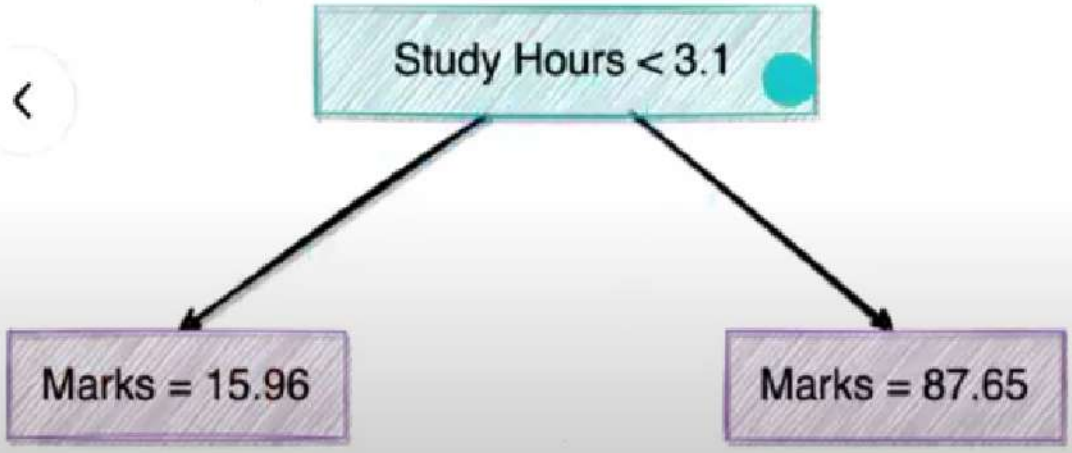

Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko
@nerchu_ko
Exam Marks
Study Hour
Source: medium
1:39 / 26:26

Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko

@nerchu_ko

6	2.9
7	3.3

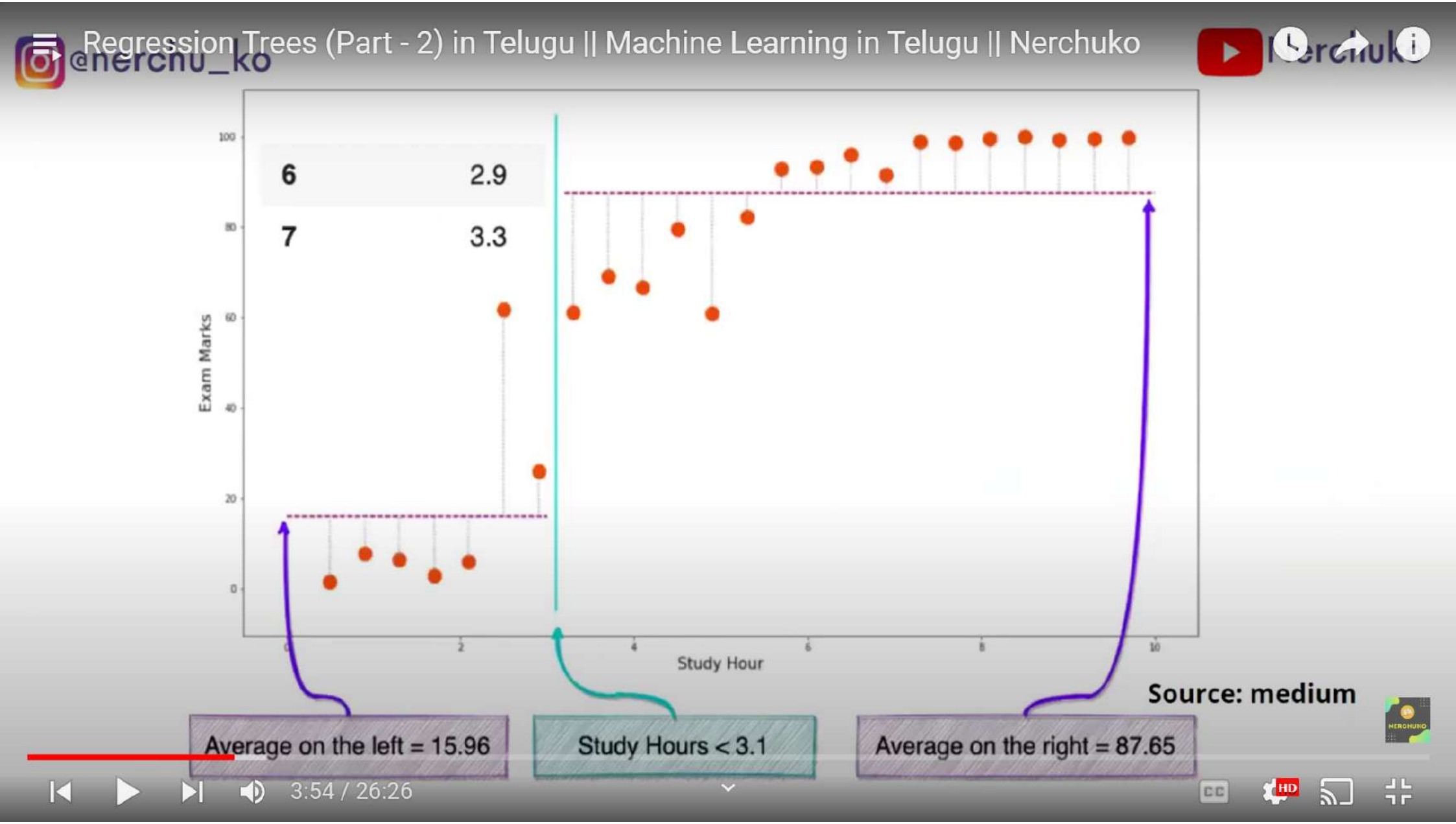
How to select the threshold value?



The Sum of Squares of residual error should be small for the decision node split

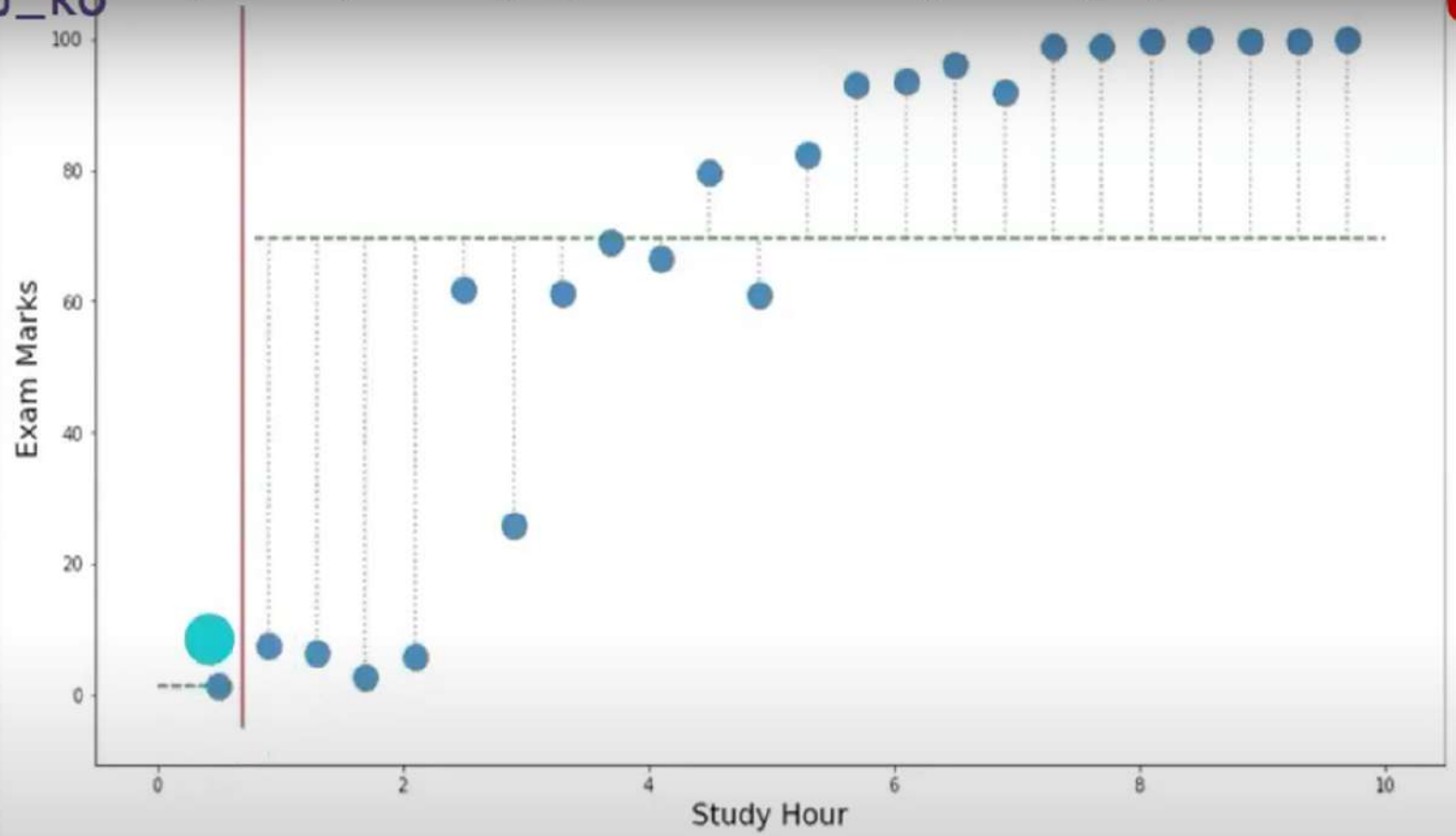
Source: medium

2:14 / 26:26


Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko
@nerchu_ko
6 2.9
7 3.3
Exam Marks
Study Hour
Source: medium
Average on the left = 15.96
Study Hours < 3.1
Average on the right = 87.65
3:54 / 26:26

Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko

@nerchu_ko



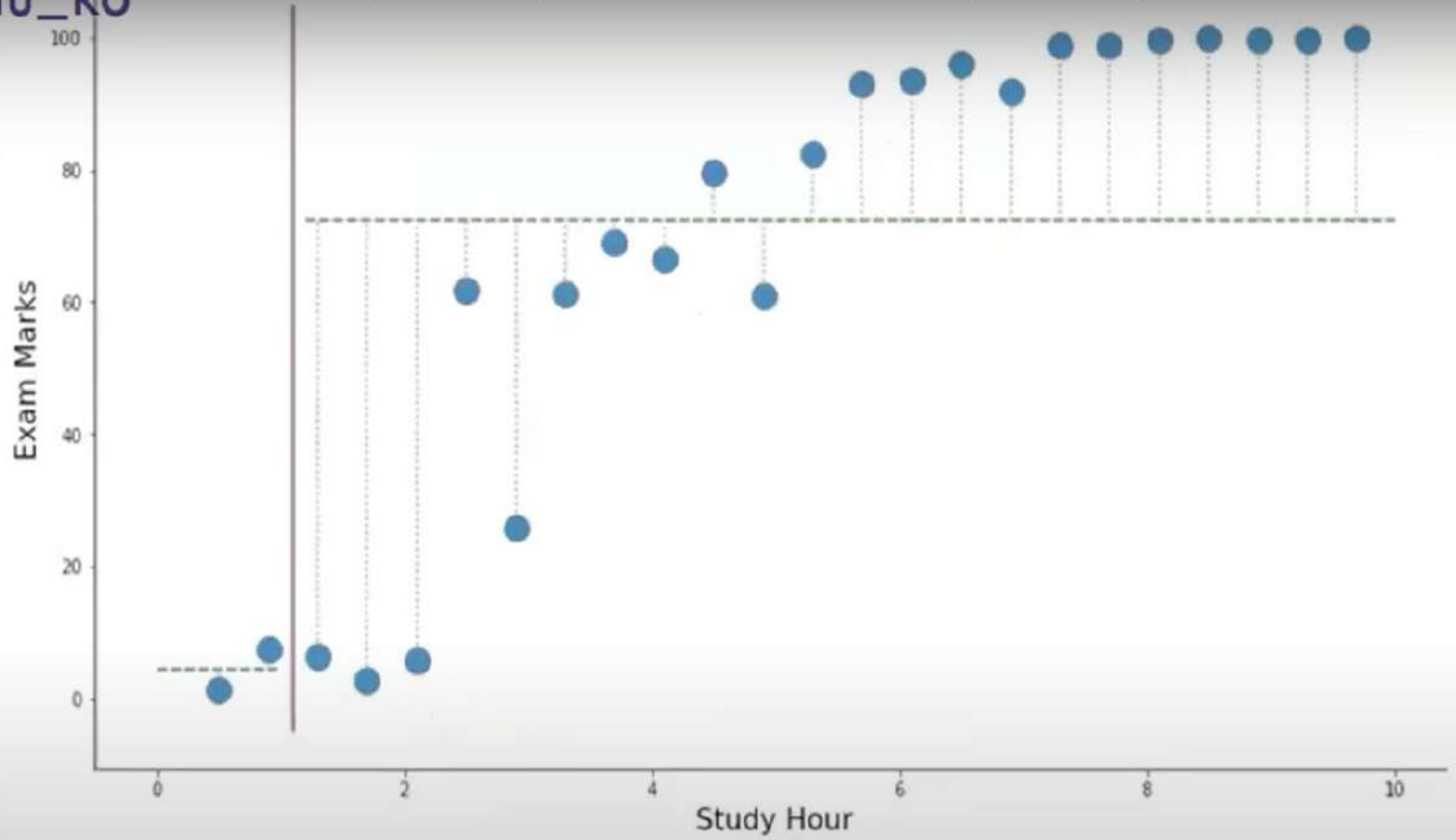

```
Split Condition: Study Hour < 0.7
Average on the left: 1.4
Average on the right: 69.57826086956521
Error on the left: 0.0
Error on the right: 27293.12
Total error: 27293.12
<function ... step>
```

Source: medium

5:39 / 26:26

Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko

@nerchu_ko



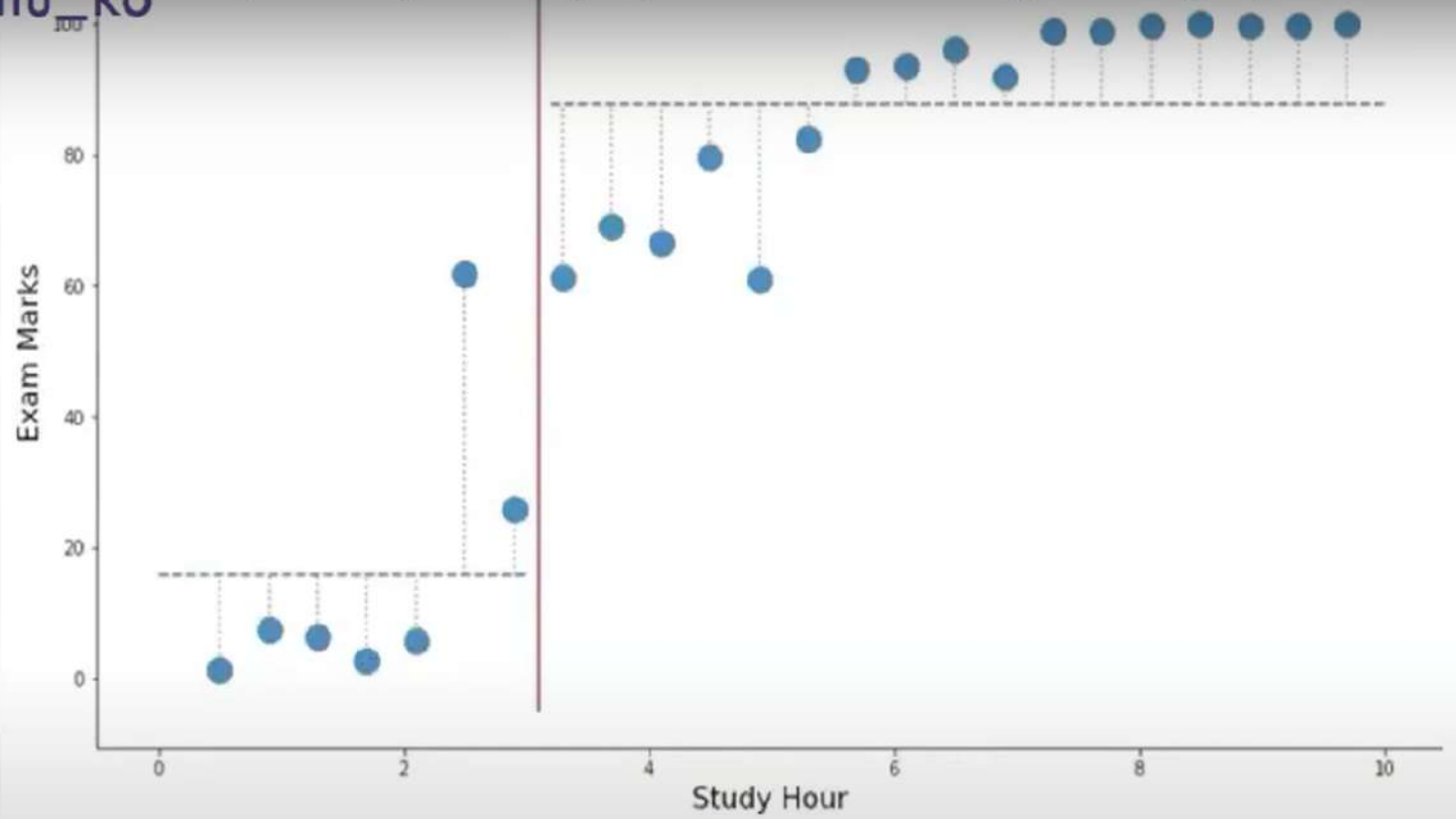

```
Split Condition: Study Hour < 1.1
Average on the left: 4.5
Average on the right: 72.39545454545454
Error on the left: 19.22
Error on the right: 23277.21
Total error: 23296.43
<function __main__.step>
```

Source: medium

7:54 / 26:26

Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko

@nerchu_ko




```
Split Condition: Study Hour < 3.1000000000000005
Average on the left: 15.957142857142859
Average on the right: 87.6470588235294
Error on the left: 2838.62
Error on the right: 3426.02
Total error: 6264.639999999999
<function __main__.step>
```

Source: medium

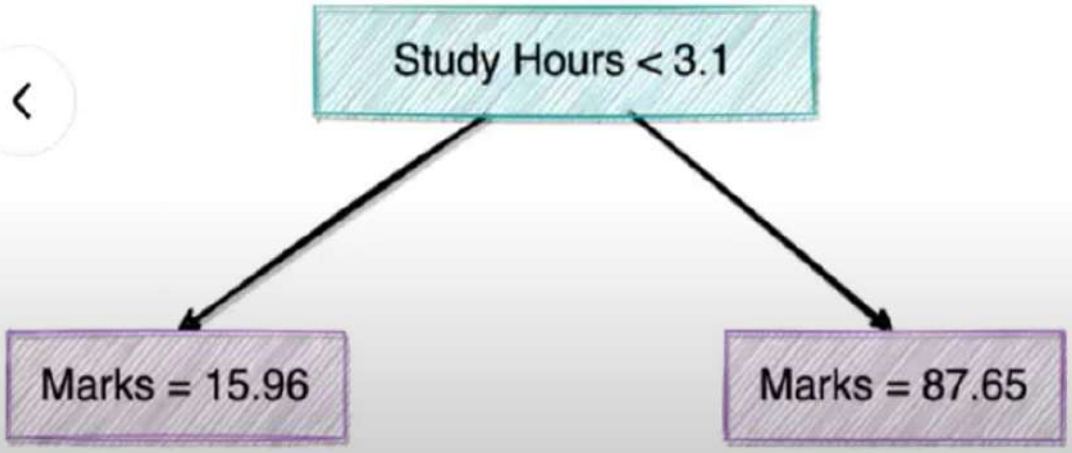
8:44 / 26:26

Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko

@nerchu_ko

6	2.9
7	3.3

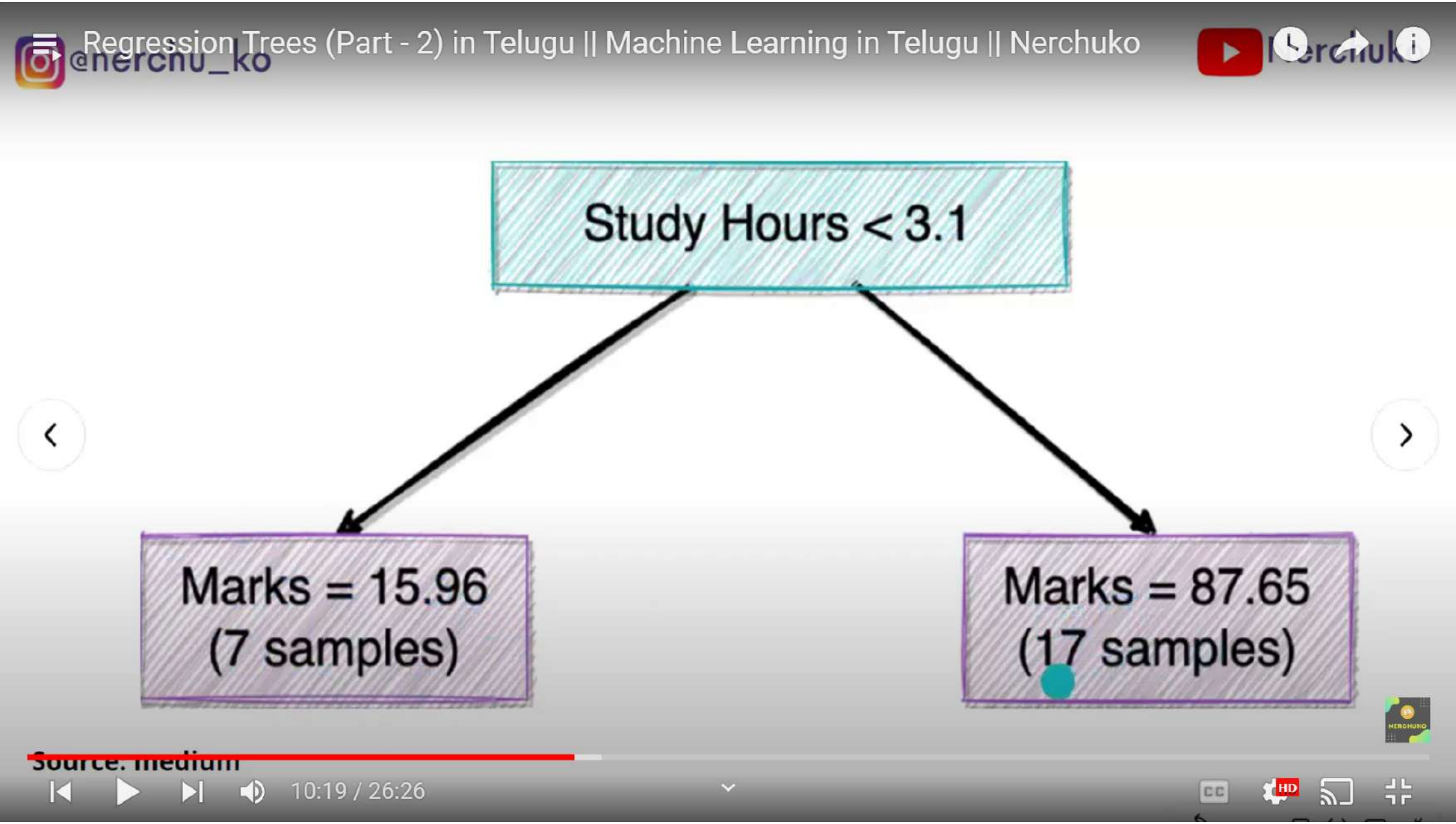
How to select the threshold value?

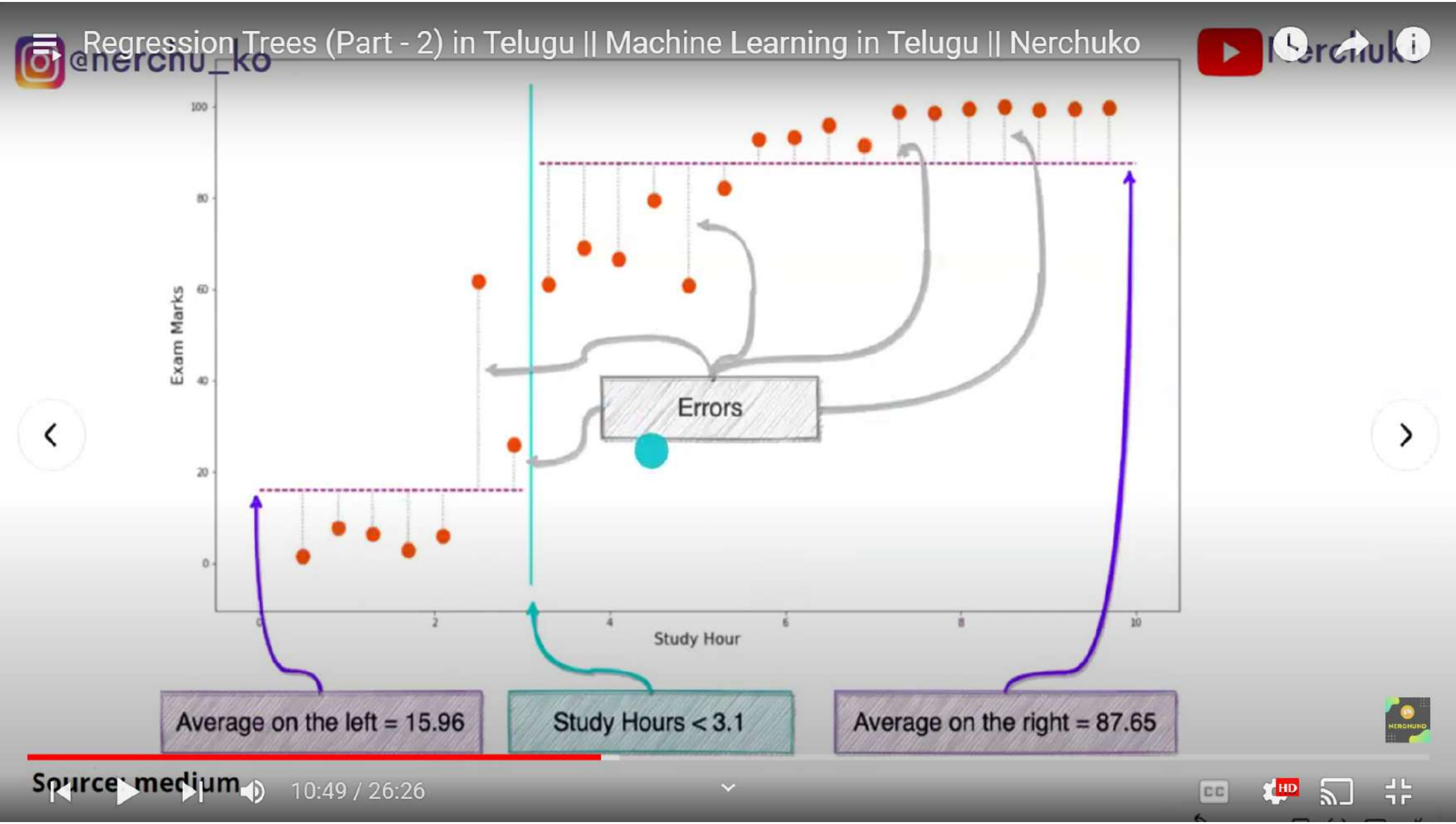


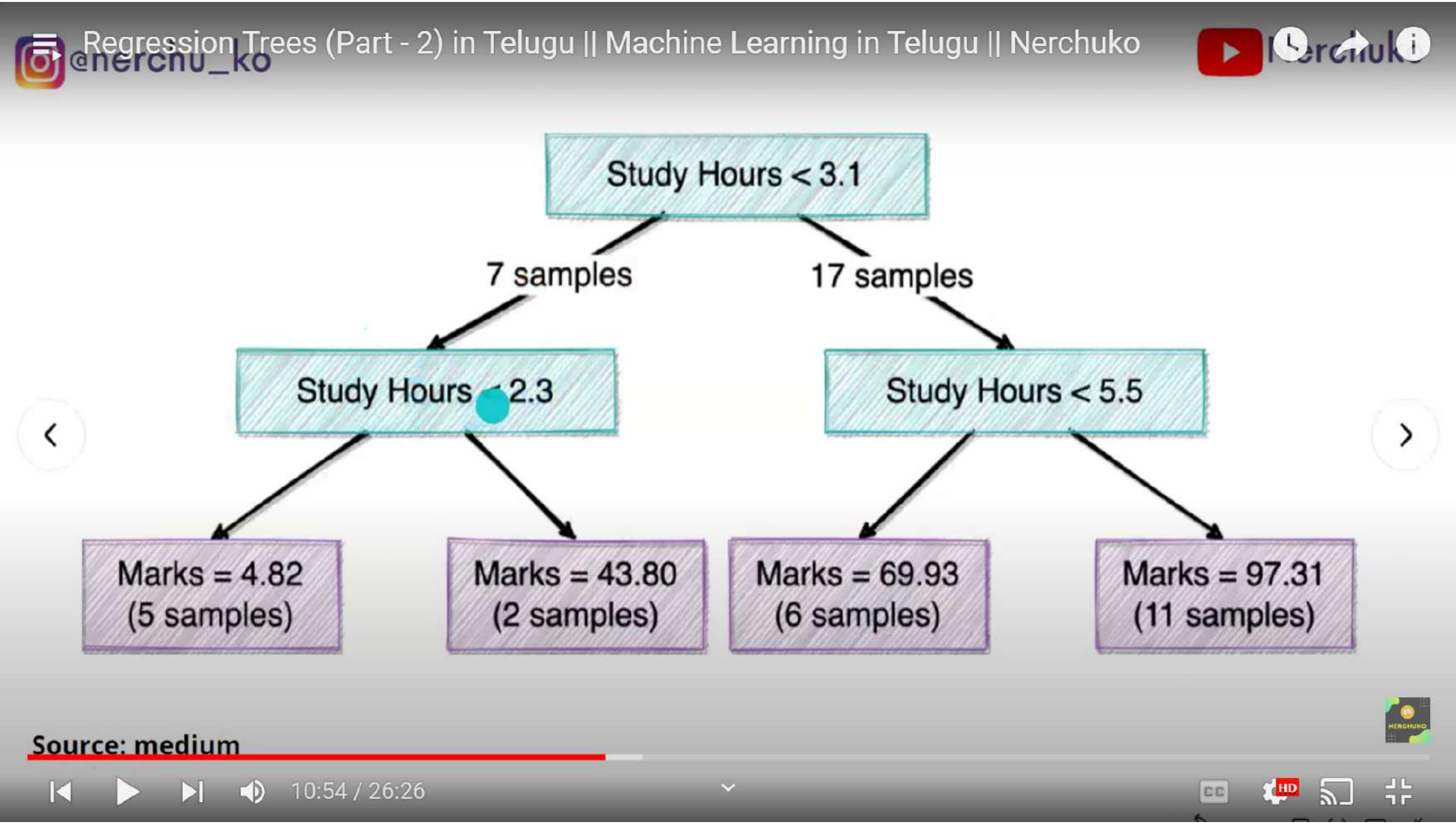
The Sum of Squares of residual error should be small for the decision node split

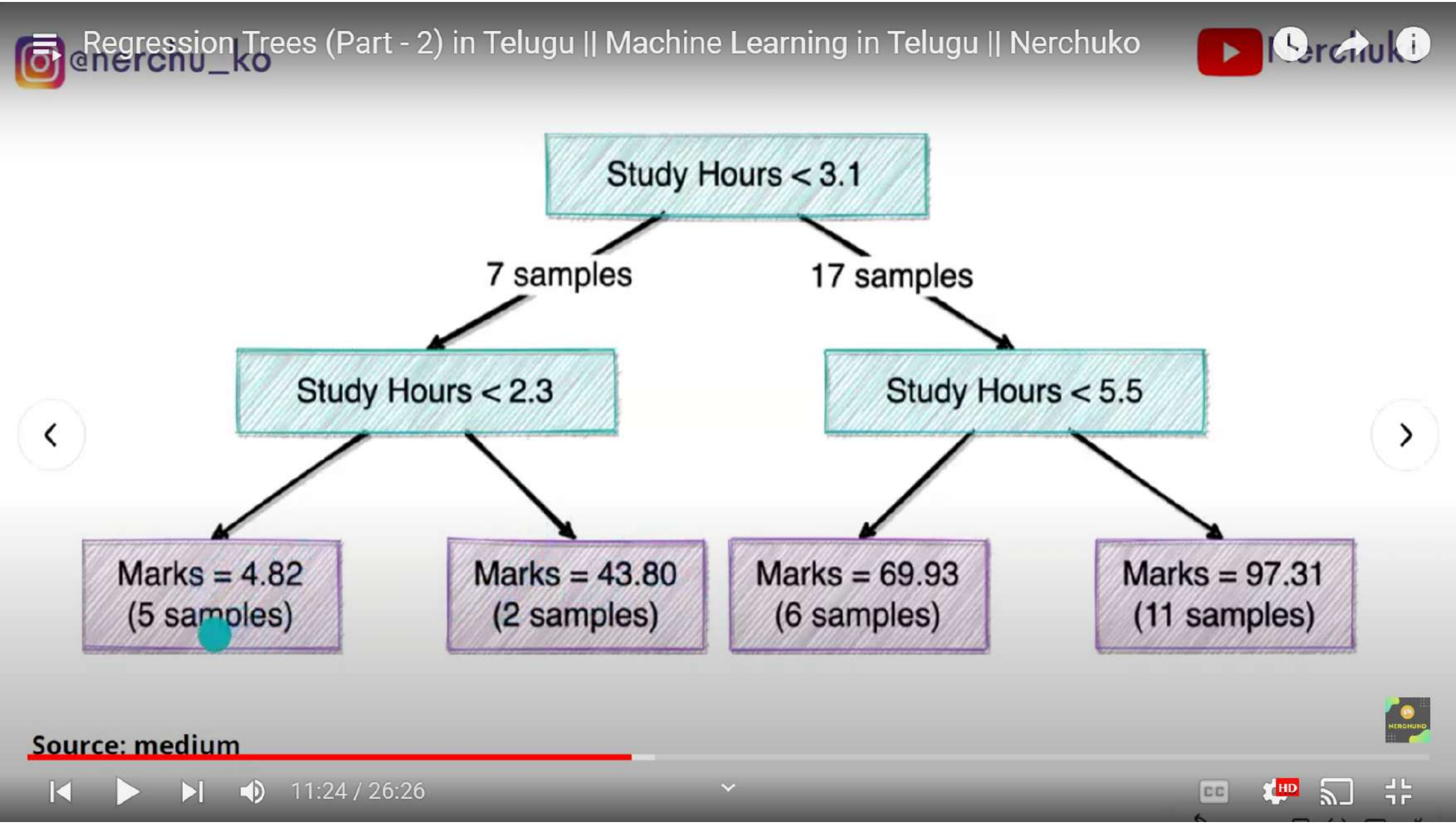
Source: medium

9:14 / 26:26


Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko
@nerchu_ko
Study Hours < 3.1
Marks = 15.96
(7 samples)
Marks = 87.65
(17 samples)
Source: Medium
10:19 / 26:26


Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko
@nerchu_ko
Exam Marks
Study Hour
Errors
Average on the left = 15.96
Study Hours < 3.1
Average on the right = 87.65
Source: medium
10:49 / 26:26

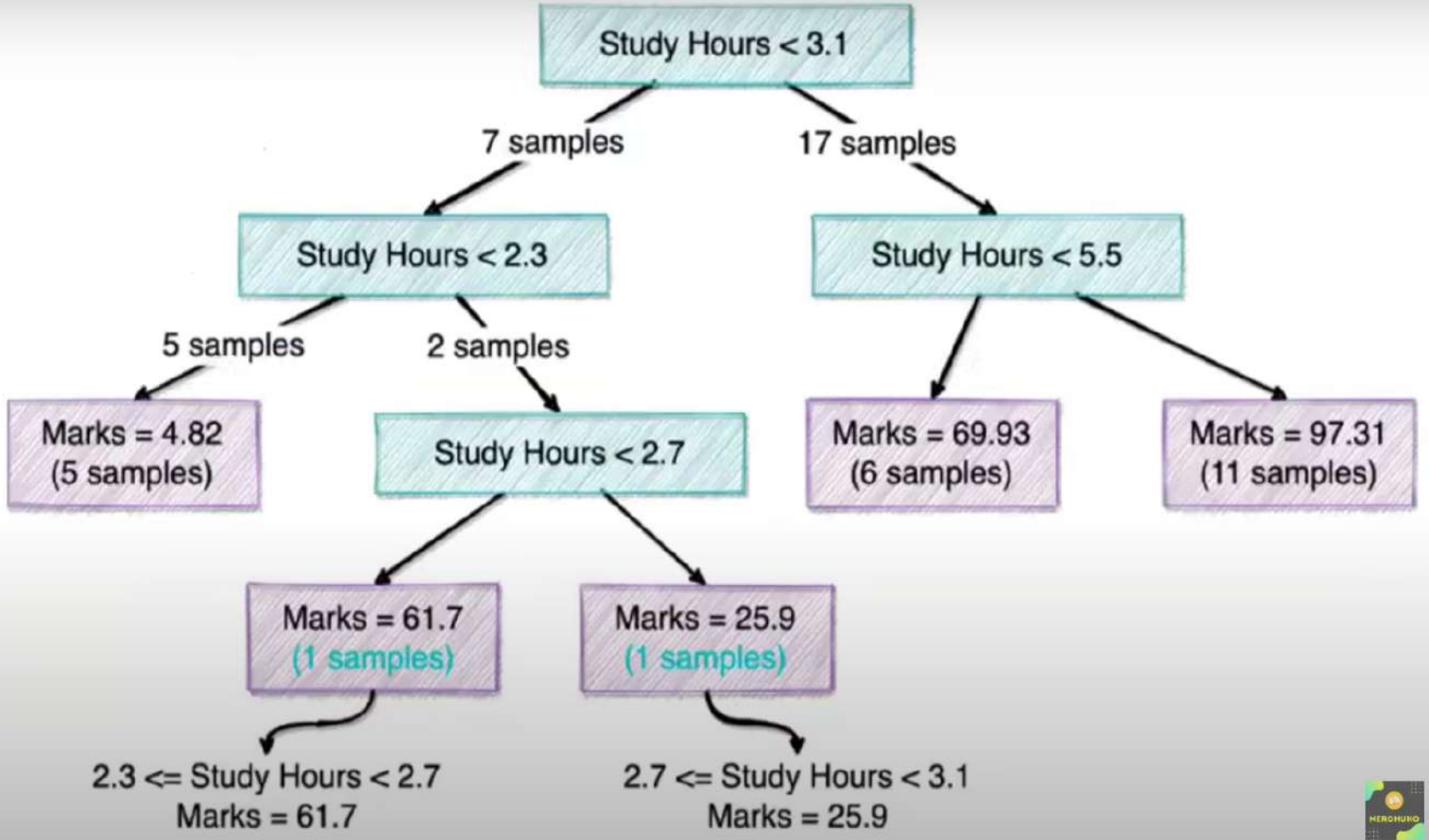

Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko
@nerchu_ko
Study Hours < 3.1
7 samples
17 samples
Study Hours < 2.3
Study Hours < 5.5
Marks = 4.82
(5 samples)
Marks = 43.80
(2 samples)
Marks = 69.93
(6 samples)
Marks = 97.31
(11 samples)
Source: medium
10:54 / 26:26


Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko
@nerchu_ko
Study Hours < 3.1
7 samples
17 samples
Study Hours < 2.3
Study Hours < 5.5
Marks = 4.82
(5 samples)
Marks = 43.80
(2 samples)
Marks = 69.93
(6 samples)
Marks = 97.31
(11 samples)
Source: medium
11:24 / 26:26

Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko

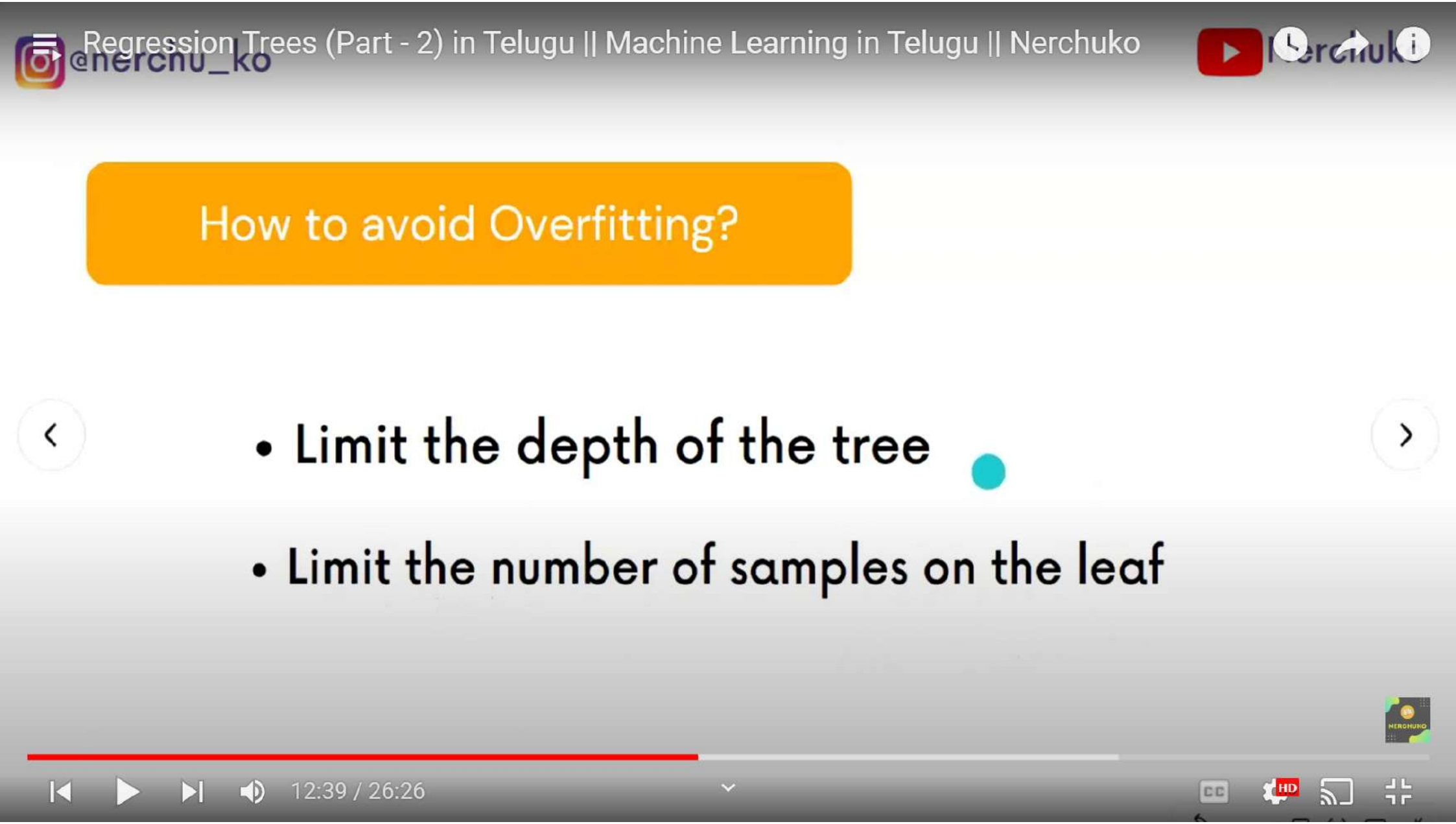
@nerchu_ko

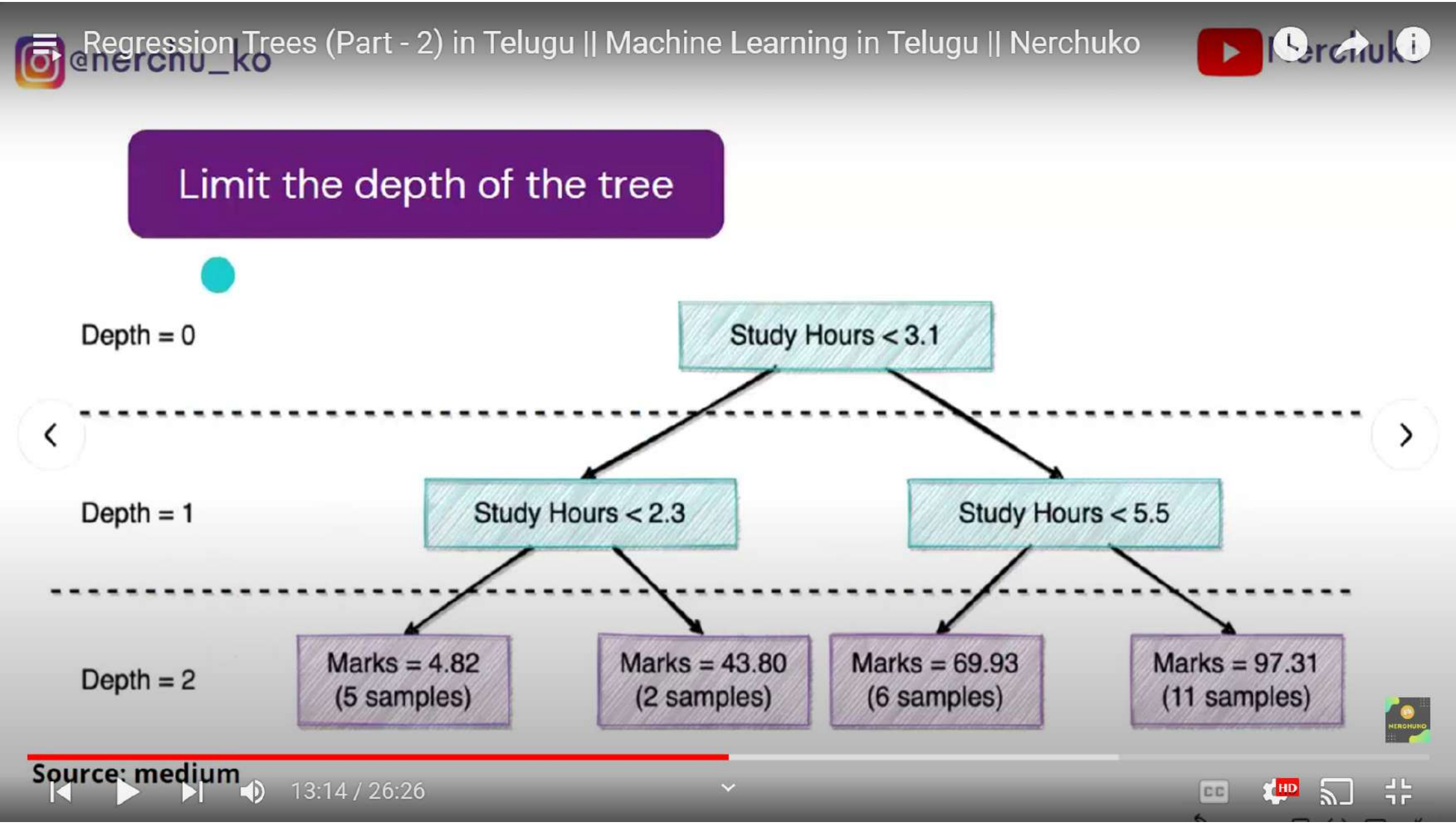
Problem of Overfitting

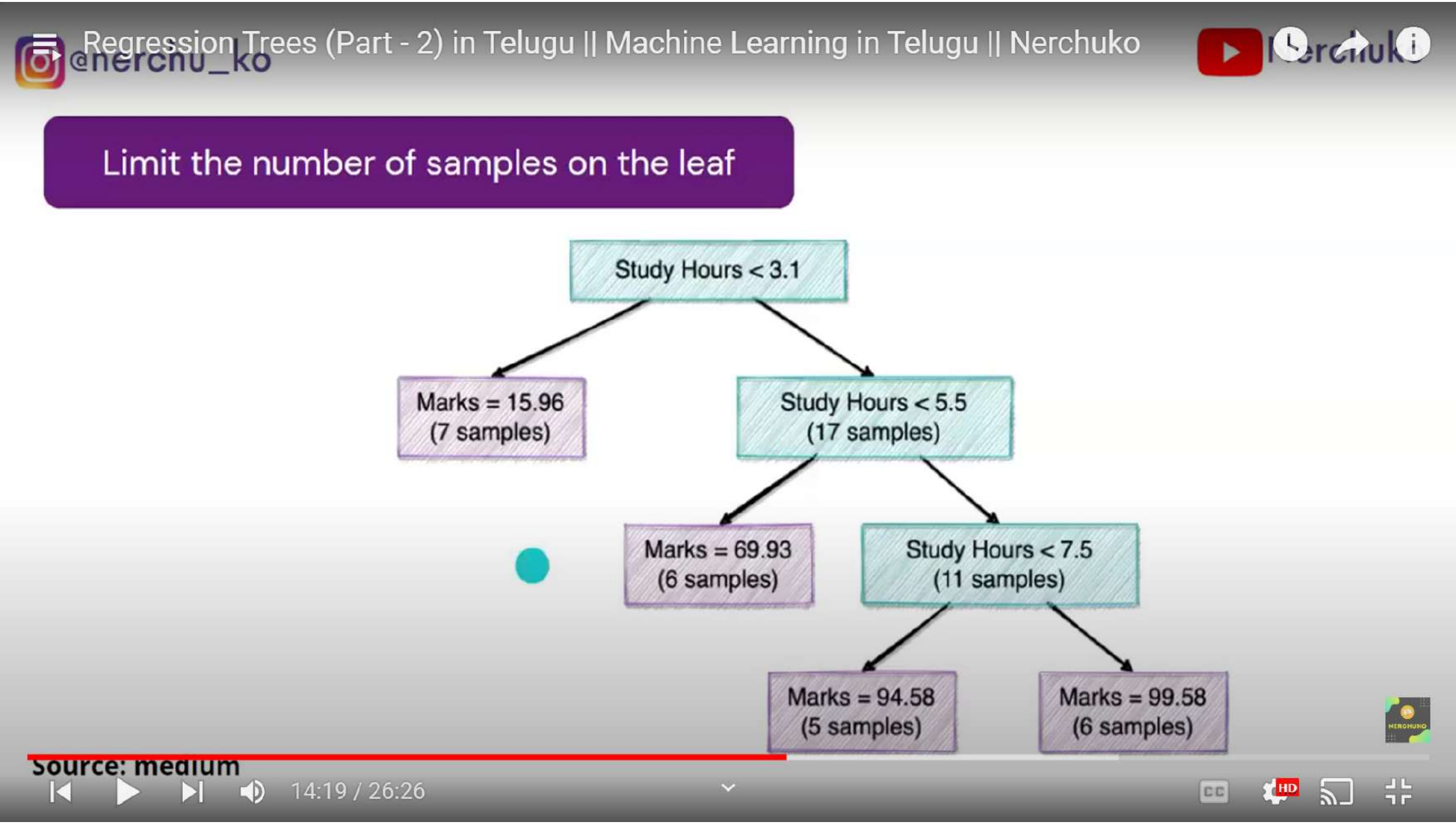


Source: medium

11:44 / 26:26


Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko
@nerchu_ko
How to avoid Overfitting?
• Limit the depth of the tree
• Limit the number of samples on the leaf
12:39 / 26:26


Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko
@nerchu_ko
Limit the depth of the tree
Depth = 0
Study Hours < 3.1
Depth = 1
Study Hours < 2.3
Study Hours < 5.5
Depth = 2
Marks = 4.82
(5 samples)
Marks = 43.80
(2 samples)
Marks = 69.93
(6 samples)
Marks = 97.31
(11 samples)
Source: medium
13:14 / 26:26


Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko
@nerchu_ko
Limit the number of samples on the leaf
Study Hours < 3.1
Marks = 15.96
(7 samples)
Study Hours < 5.5
(17 samples)
Marks = 69.93
(6 samples)
Study Hours < 7.5
(11 samples)
Marks = 94.58
(5 samples)
Marks = 99.58
(6 samples)
Source: medium
14:19 / 26:26

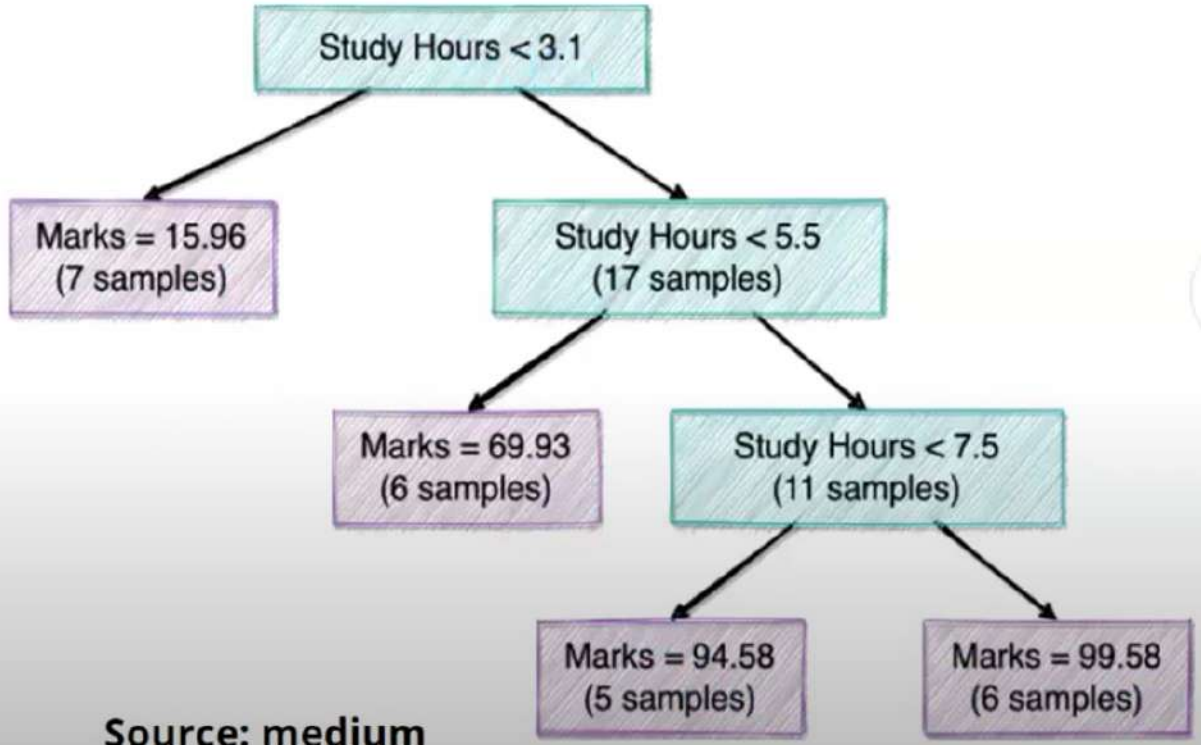
Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko

@nerchu_ko

Limit the number of samples on the leaf

	Study Hours	Marks
0	0.5	1.4
1	0.9	7.6
2	1.3	6.4
3	1.7	2.8
4	2.1	5.9
5	2.5	61.7
6	2.9	25.9
7	3.3	61.1
8	3.7	69.1
9	4.1	66.6
10	4.5	79.5
11	4.9	61.0

	Study Hours	Marks
12	5.3	82.3
13	5.7	92.9
14	6.1	93.4
15	6.5	96.0
16	6.9	91.7
17	7.3	98.9
18	7.7	98.8
19	8.1	99.7
20	8.5	100.0
21	8.9	99.5
22	9.3	99.6
23	9.7	99.9

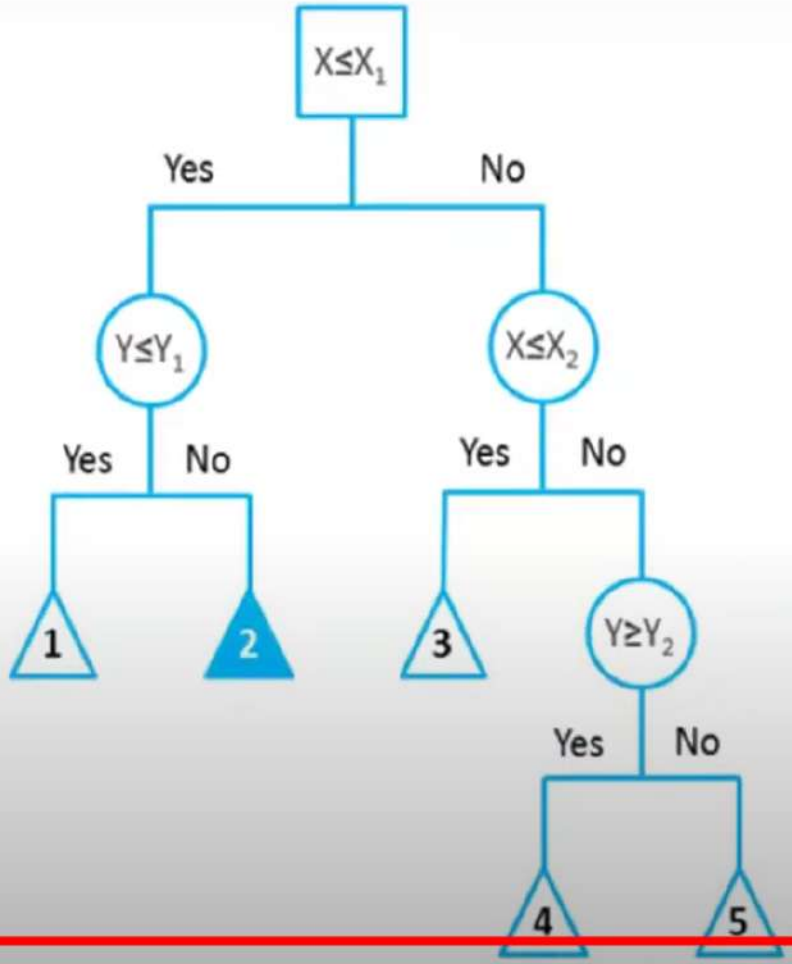


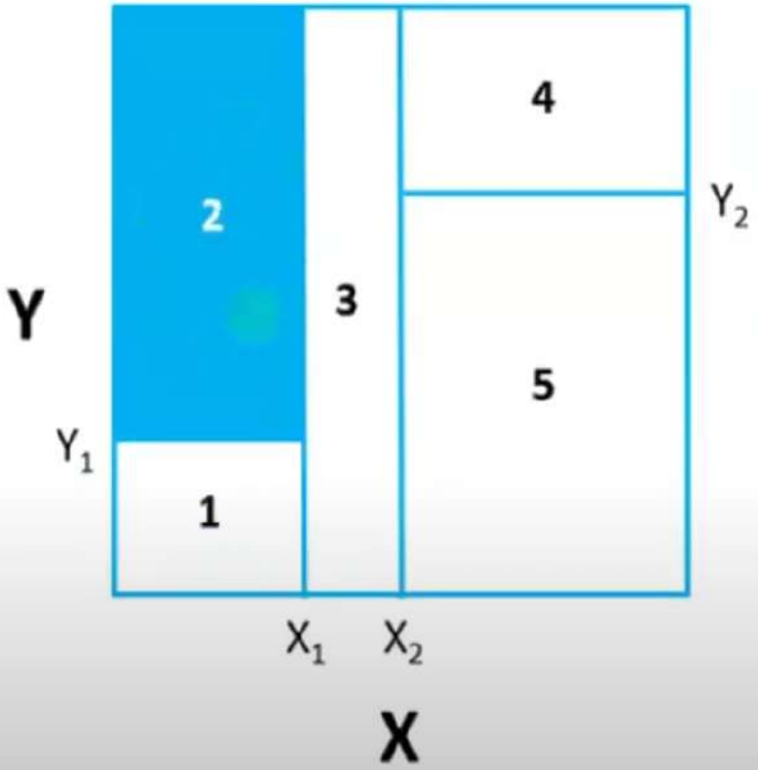
Source: medium

15:29 / 26:26

Regression Trees (Part - 2) in Telugu || Machine Learning in Telugu || Nerchuko

@nerchu_ko





19:29 / 26:26