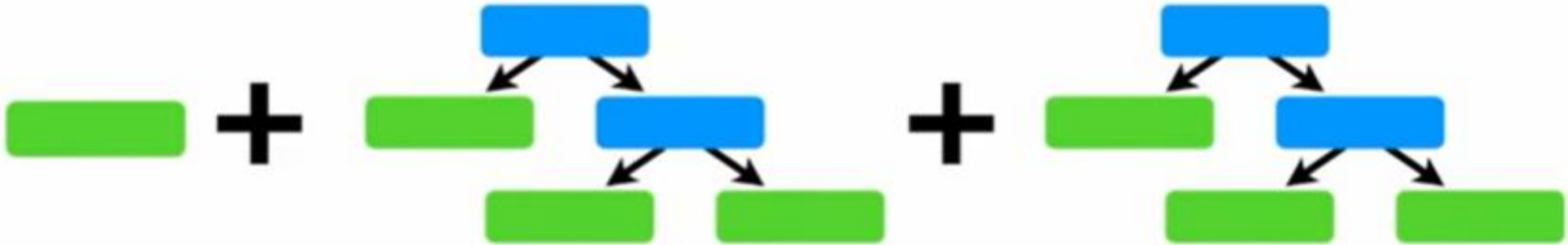
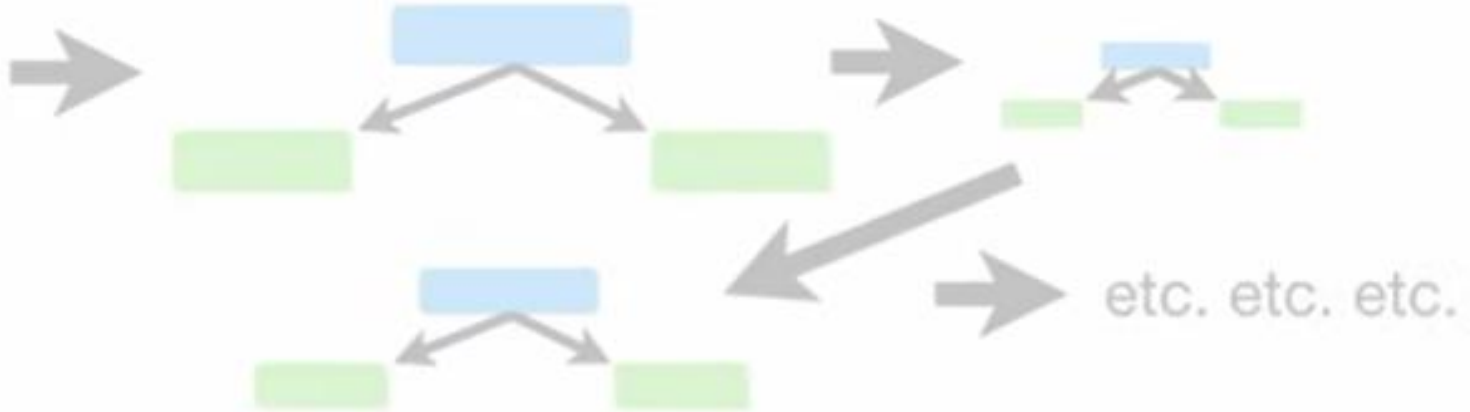
Gradient Boost Part 1...



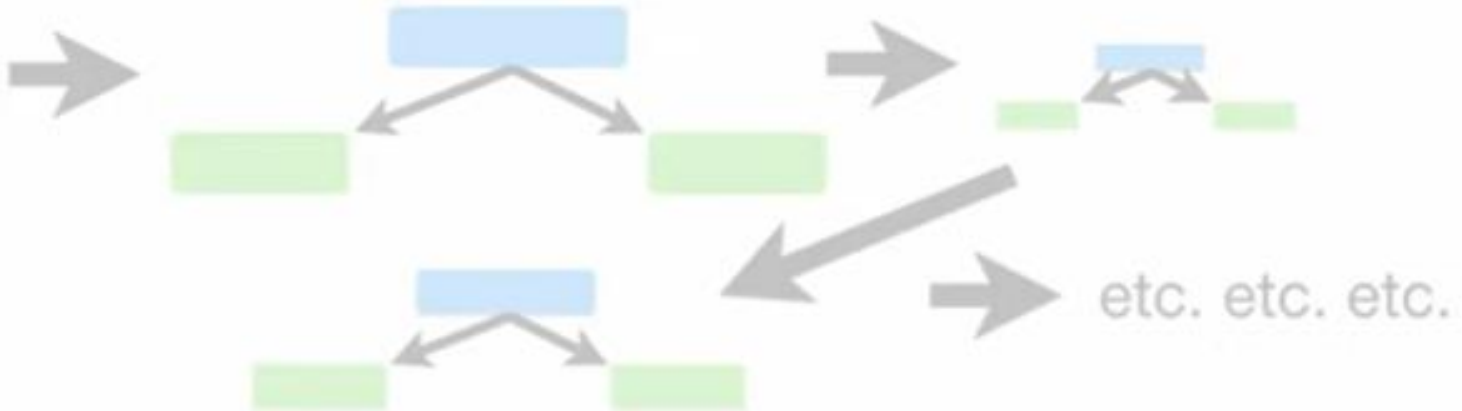
...Regression Main Ideas!!!

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc…	etc…	etc…	etc…



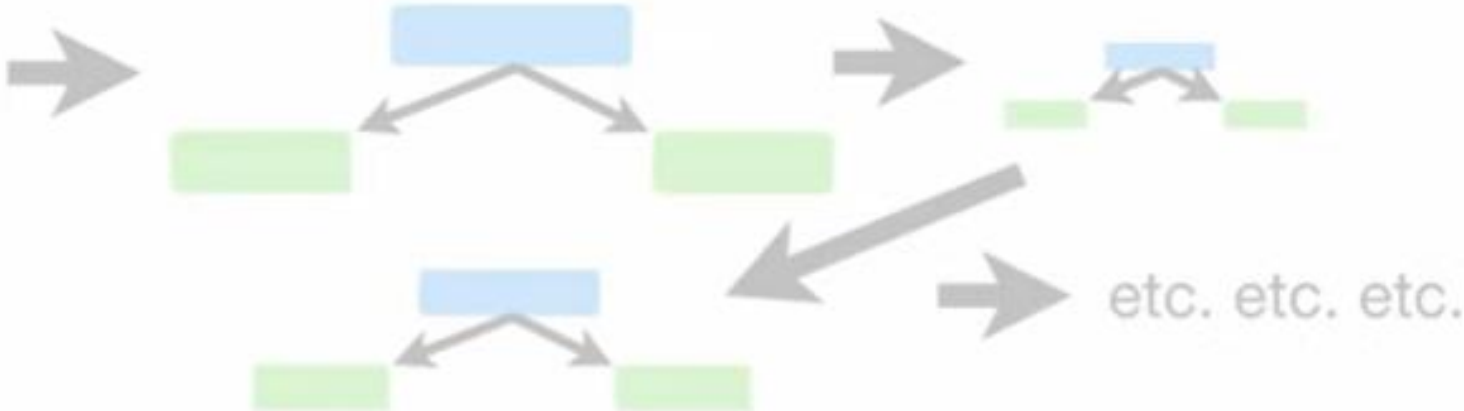
In contrast, **Gradient Boost** starts by making a single leaf, instead of a tree or stump.

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc…	etc…	etc…	etc…



This leaf represents an initial guess for the **Weights** of all of the samples.

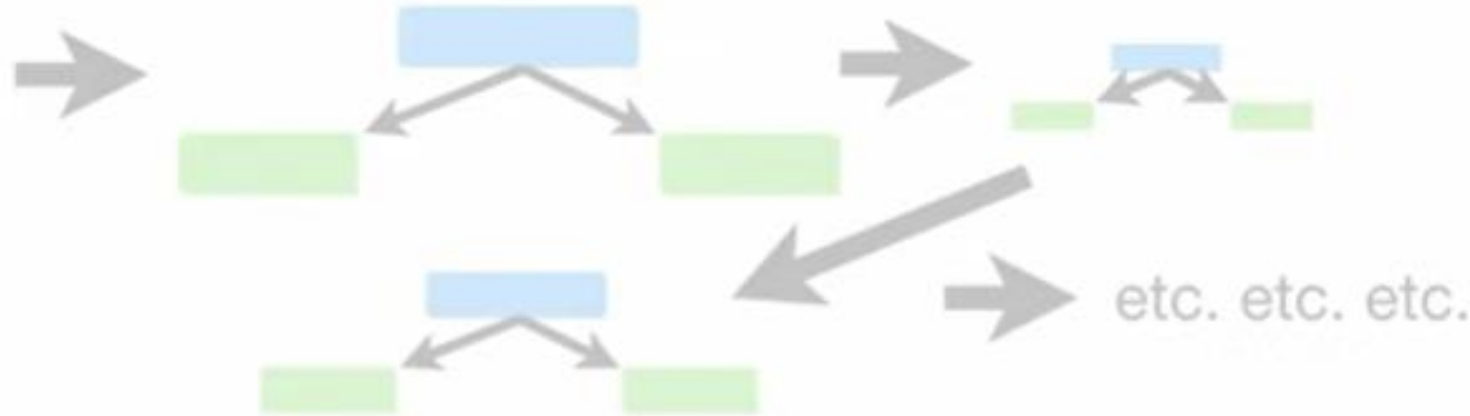
Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc…	etc…	etc…	etc…

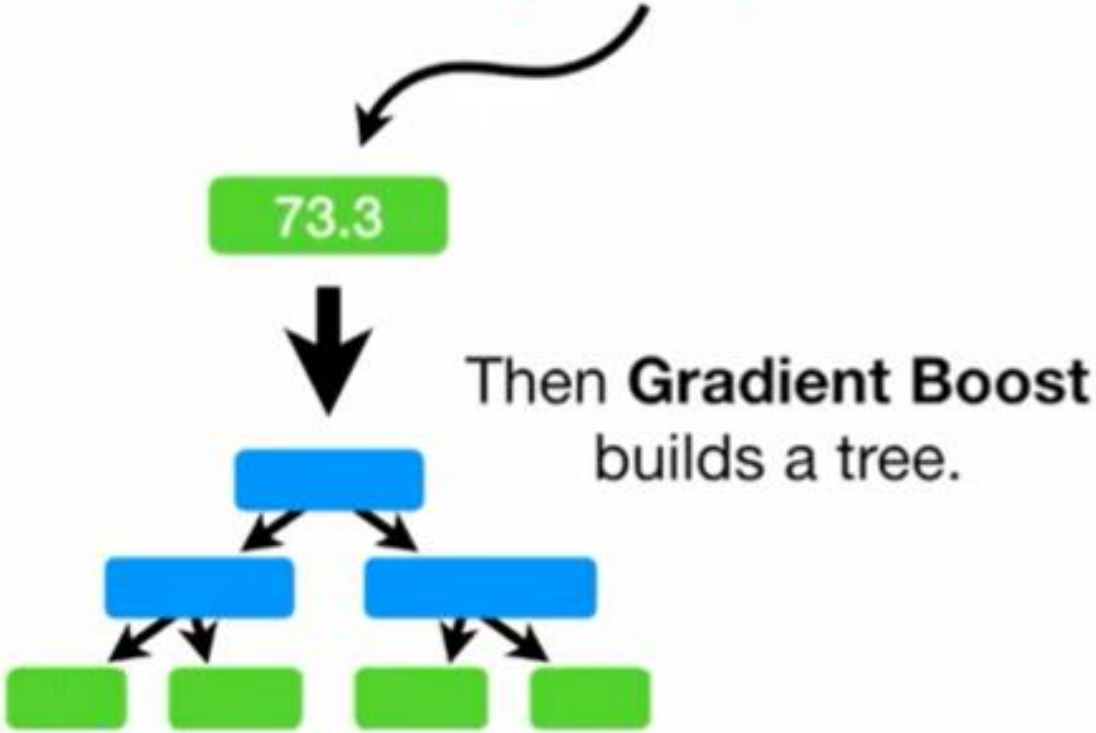


73.3

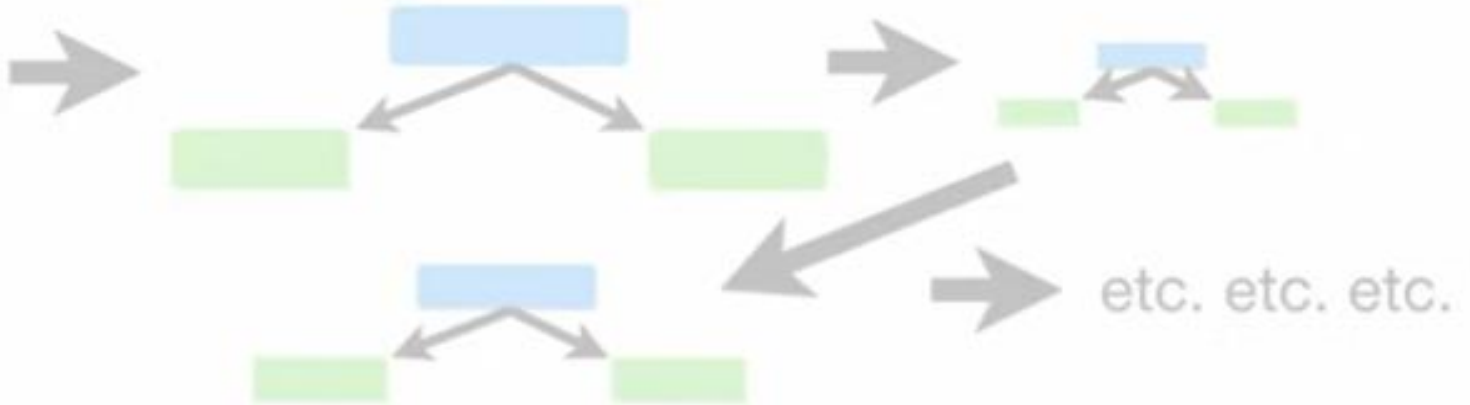
When trying to **Predict** a continuous value like **Weight**, the first guess is the the average value.

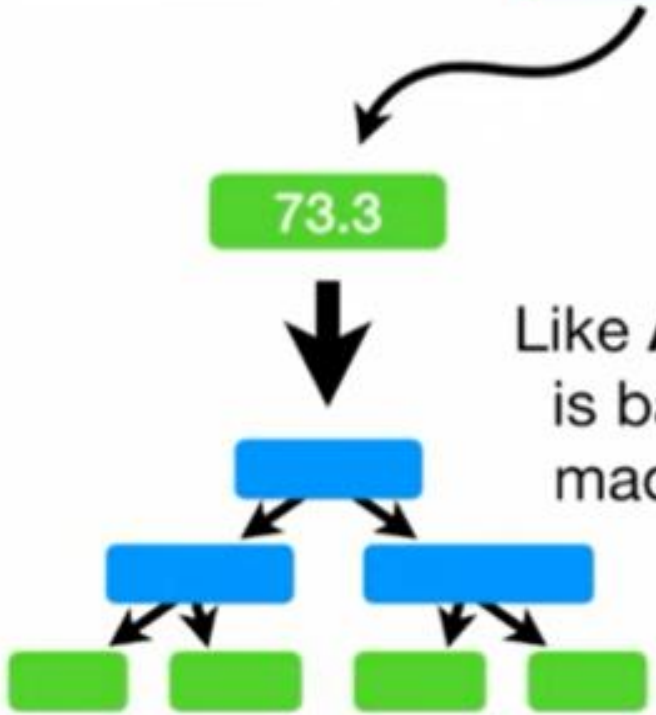
Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc…	etc…	etc…	etc…





Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc…	etc…	etc…	etc…

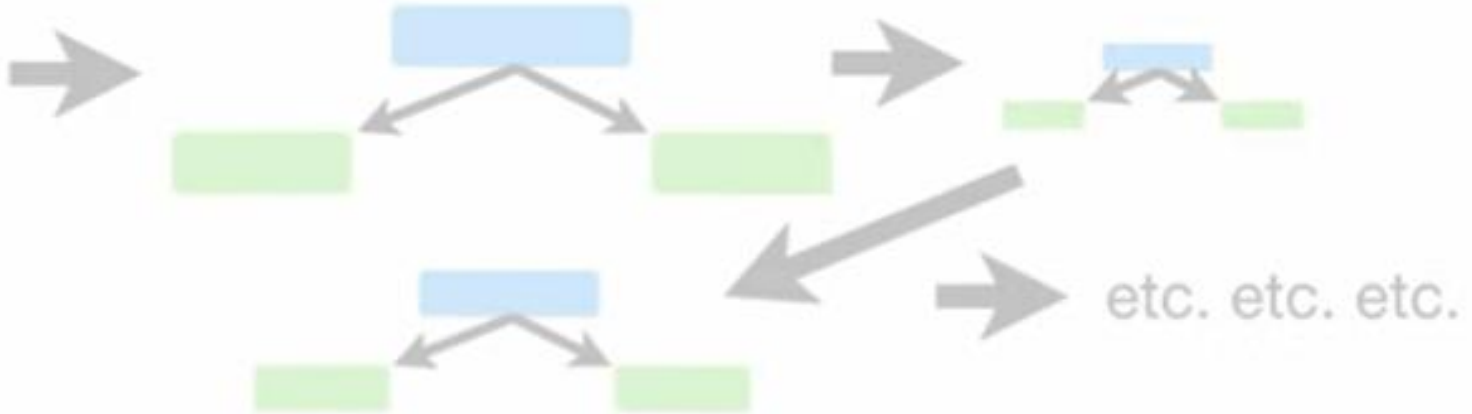




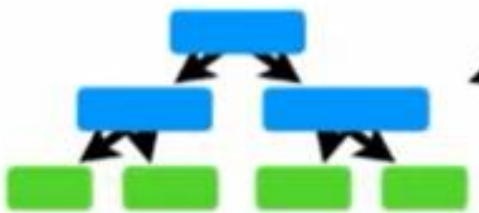
Like **AdaBoost**, this tree is based on the errors made by the previous tree…

…but unlike **AdaBoost**, this tree is usually larger than a stump.

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc…	etc…	etc…	etc…

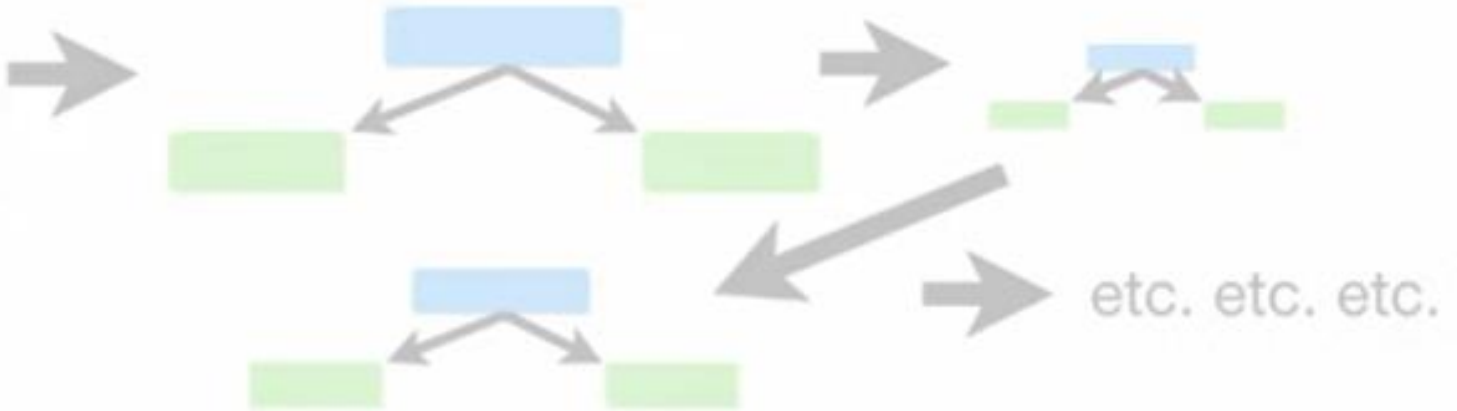


73.3



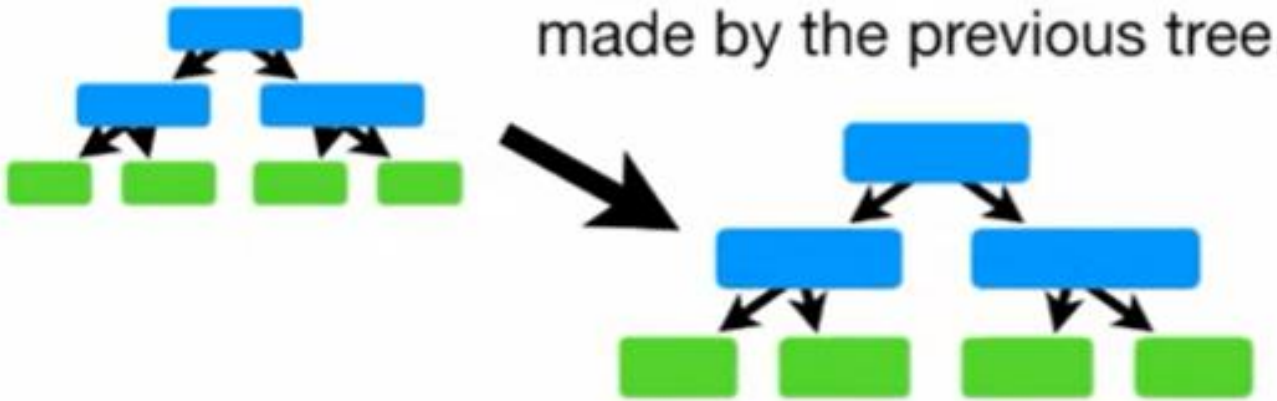
Also like **AdaBoost**, **Gradient Boost** scales the trees. However, **Gradient Boost** scales all trees by the same amount.

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc…	etc…	etc…	etc…

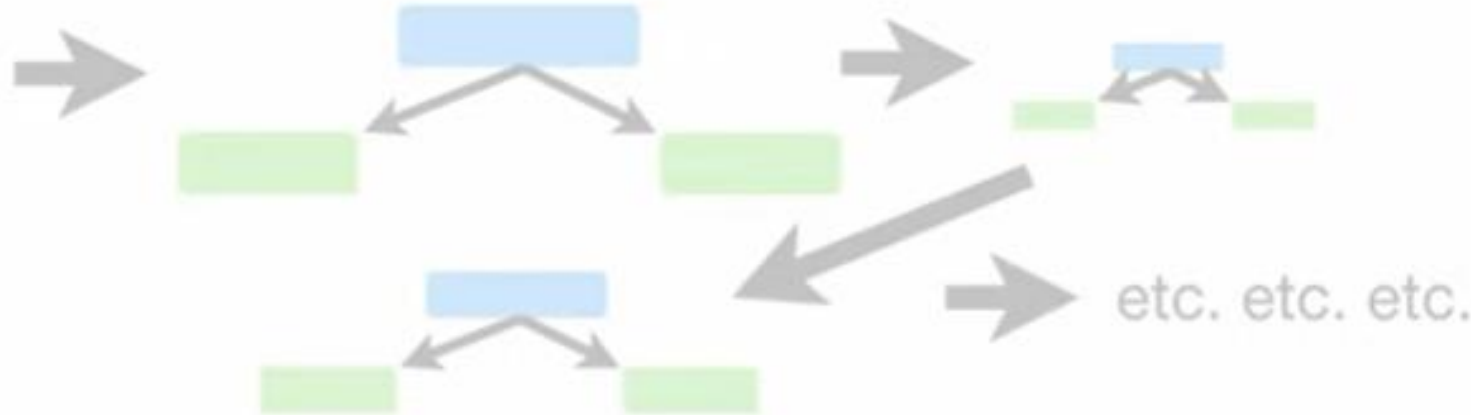


73.3

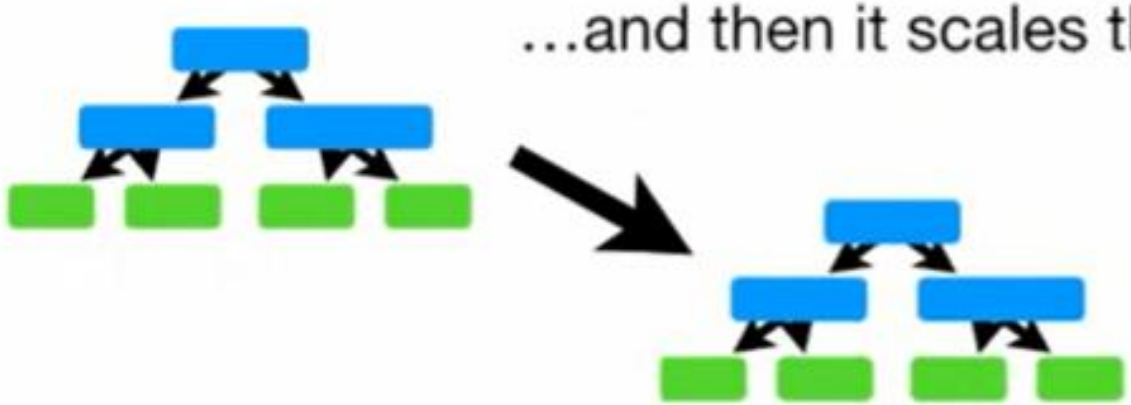
Then **Gradient Boost** builds another tree based on the errors made by the previous tree…



Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc…	etc…	etc…	etc…

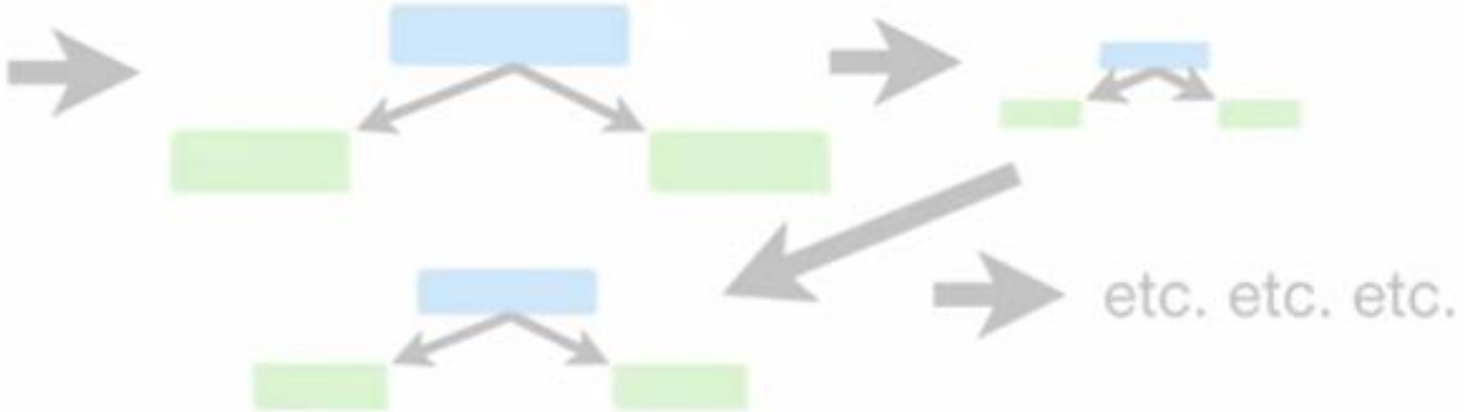


73.3



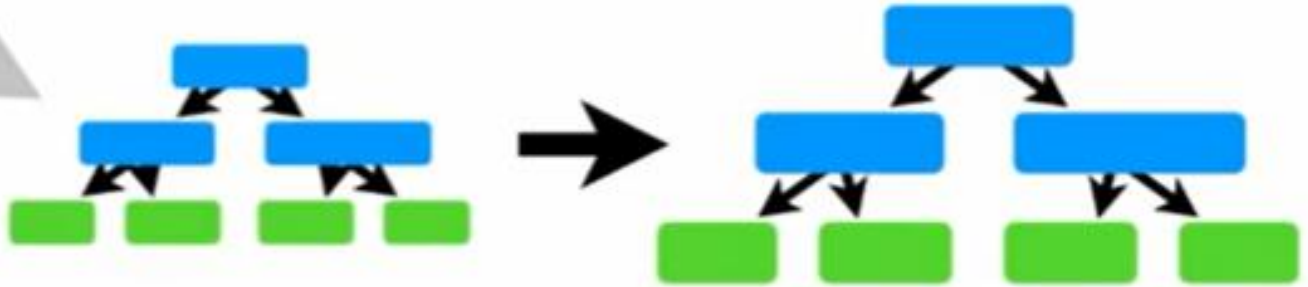
…and then it scales the tree…

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc…	etc…	etc…	etc…

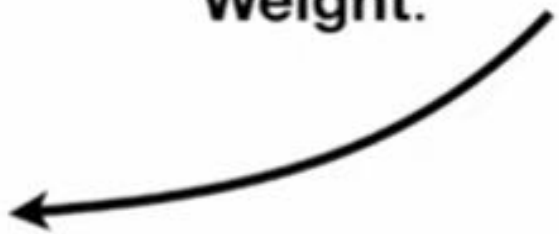


73.3

…and **Gradient Boost** continues to build trees in this fashion until it has made the number of trees you asked for, or additional trees fail to improve the fit.



…let's see how the most common **Gradient Boost** configuration would use this **Training Data** to **Predict Weight**.



Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

Average Weight

71.2

The first thing we do is calculate the average **Weight.**

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

Average Weight

71.2

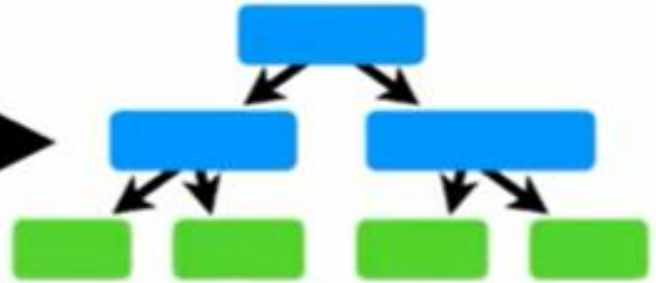
Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

In other words, if we stopped right now, we would predict that everyone **Weighed 71.2** kg.

However, **Gradient Boost** doesn't stop here.

Average Weight

71.2



Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

The next thing we do is build a tree based on the errors from the first tree.

Average Weight

71.2

The errors that the previous tree made are the differences between the **Observed Weights** and the **Predicted Weight**, **71.2**.

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

(Observed Weight - Predicted Weight)

Average Weight

71.2

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

So let's start by plugging in **71.2** for the **Predicted Weight**...

(**Observed Weight - 71.2**)

Average Weight

71.2

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	
1.5	Blue	Female	56	
1.8	Red	Male	73	
1.5	Green	Male	77	
1.4	Blue	Female	57	

…and save the difference, which is called a **Pseudo Residual**, in a new column.

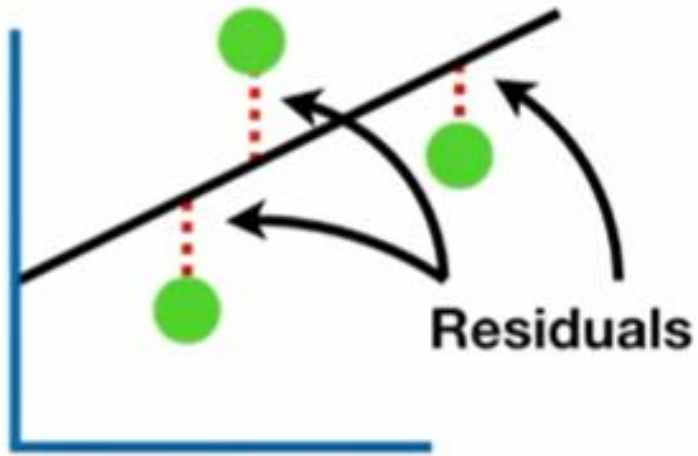
(**88** - **71.2**) = 16.8

Average Weight

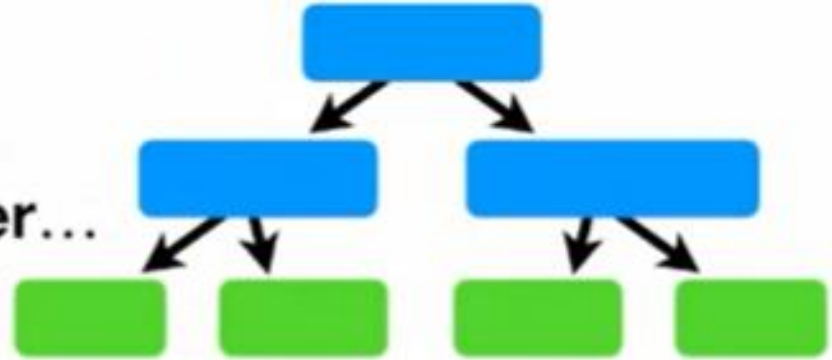
71.2

NOTE: The term **Pseudo Residual** is based on **Linear Regression**, where the difference between the **Observed** values and the **Predicted** values results in **Residuals**.

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	
1.5	Blue	Female	56	
1.8	Red	Male	73	
1.5	Green	Male	77	
1.4	Blue	Female	57	

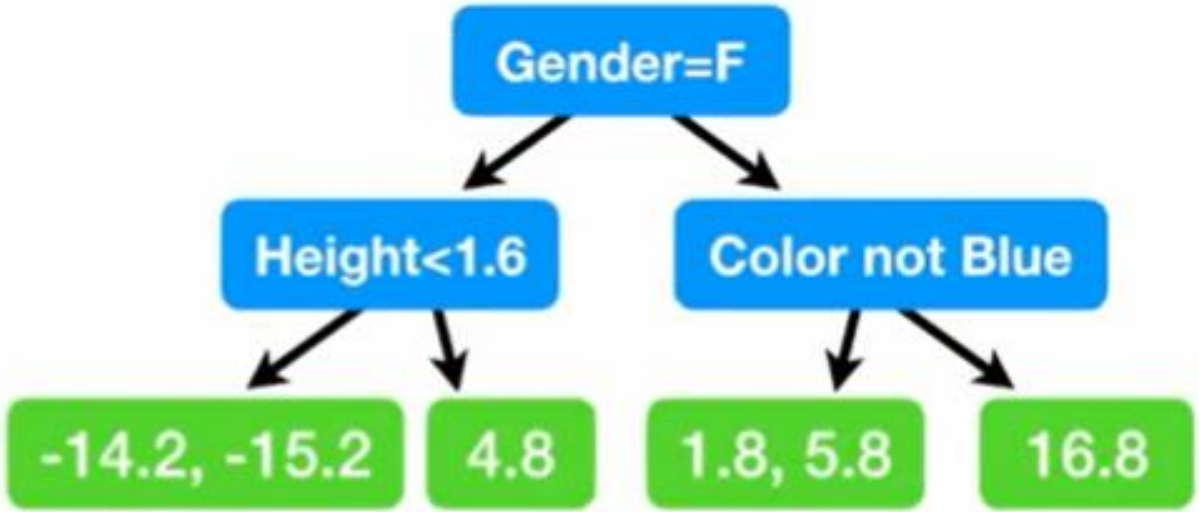


Now we will build a **Tree**, using **Height**, **Favorite Color** and **Gender**…



Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
1.5	Blue	Female	56	-15.2
1.8	Red	Male	73	1.8
1.5	Green	Male	77	5.8
1.4	Blue	Female	57	-14.2

Height (m)	Favorite Color	Gender		Residual
1.6	Blue	Male		16.8
1.6	Green	Female		4.8
1.5	Blue	Female		-15.2
1.8	Red	Male		1.8
1.5	Green	Male		5.8
1.4	Blue	Female		-14.2



So, setting aside the reason why we are building a tree to **Predict** the **Residuals** for the time being, here's the tree!

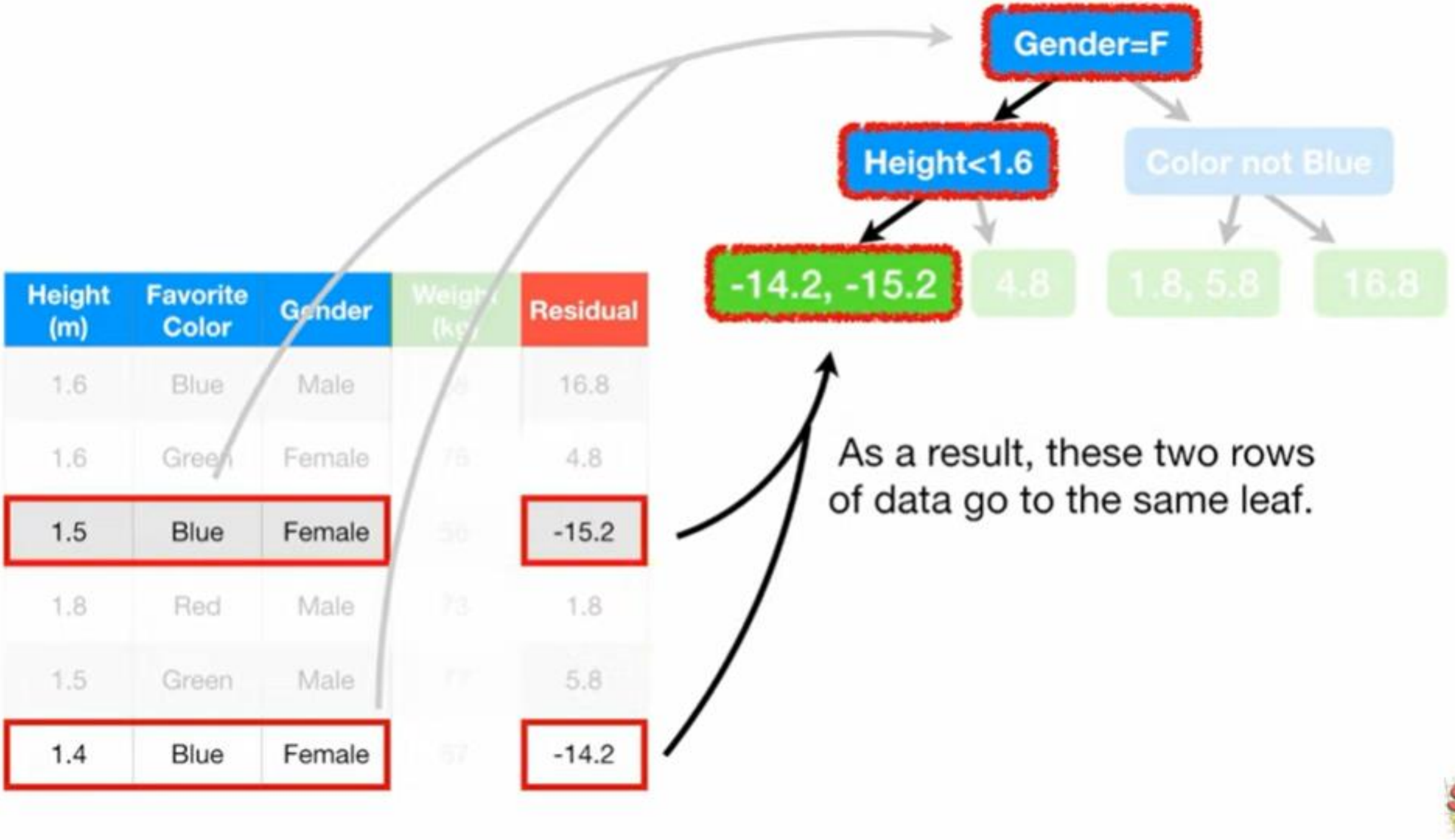
Height (m)	Favorite Color	Gender		Residual
1.6	Blue	Male		16.8
1.6	Green	Female		4.8
1.5	Blue	Female		-15.2
1.8	Red	Male		1.8
1.5	Green	Male		5.8
1.4	Blue	Female		-14.2

Remember, in this example we are only allowing up to four leaves…

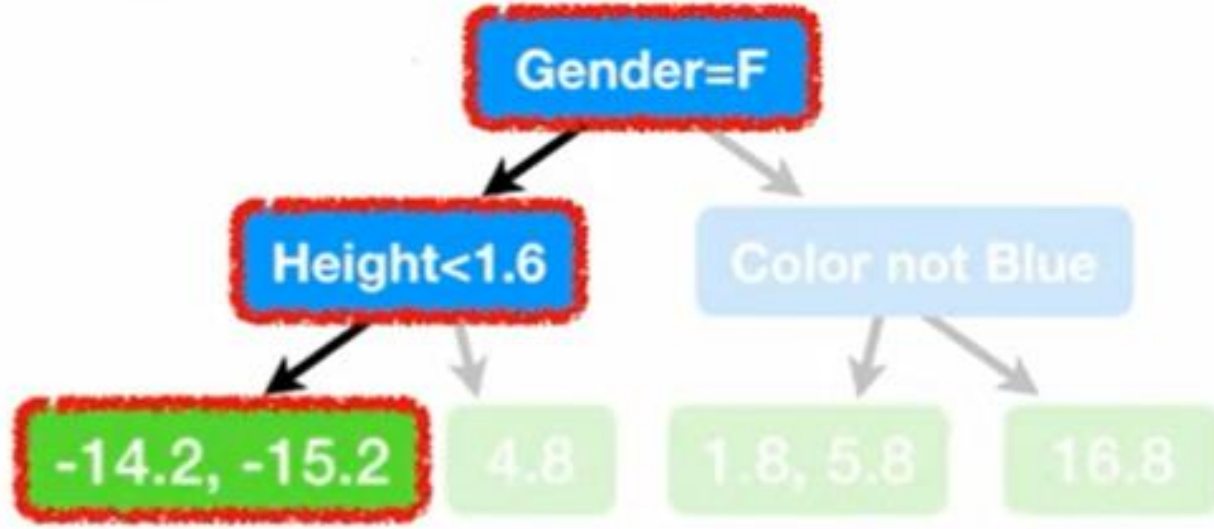
…but when using a larger dataset, it is common to allow anywhere from **8** to **32**.

Height (m)	Favorite Color	Gender	Residual
1.6	Blue	Male	16.8
1.6	Green	Female	4.8
1.5	Blue	Female	-15.2
1.8	Red	Male	1.8
1.5	Green	Male	5.8
1.4	Blue	Female	-14.2

By restricting the total number of leaves, we get fewer leaves than **Residuals**.

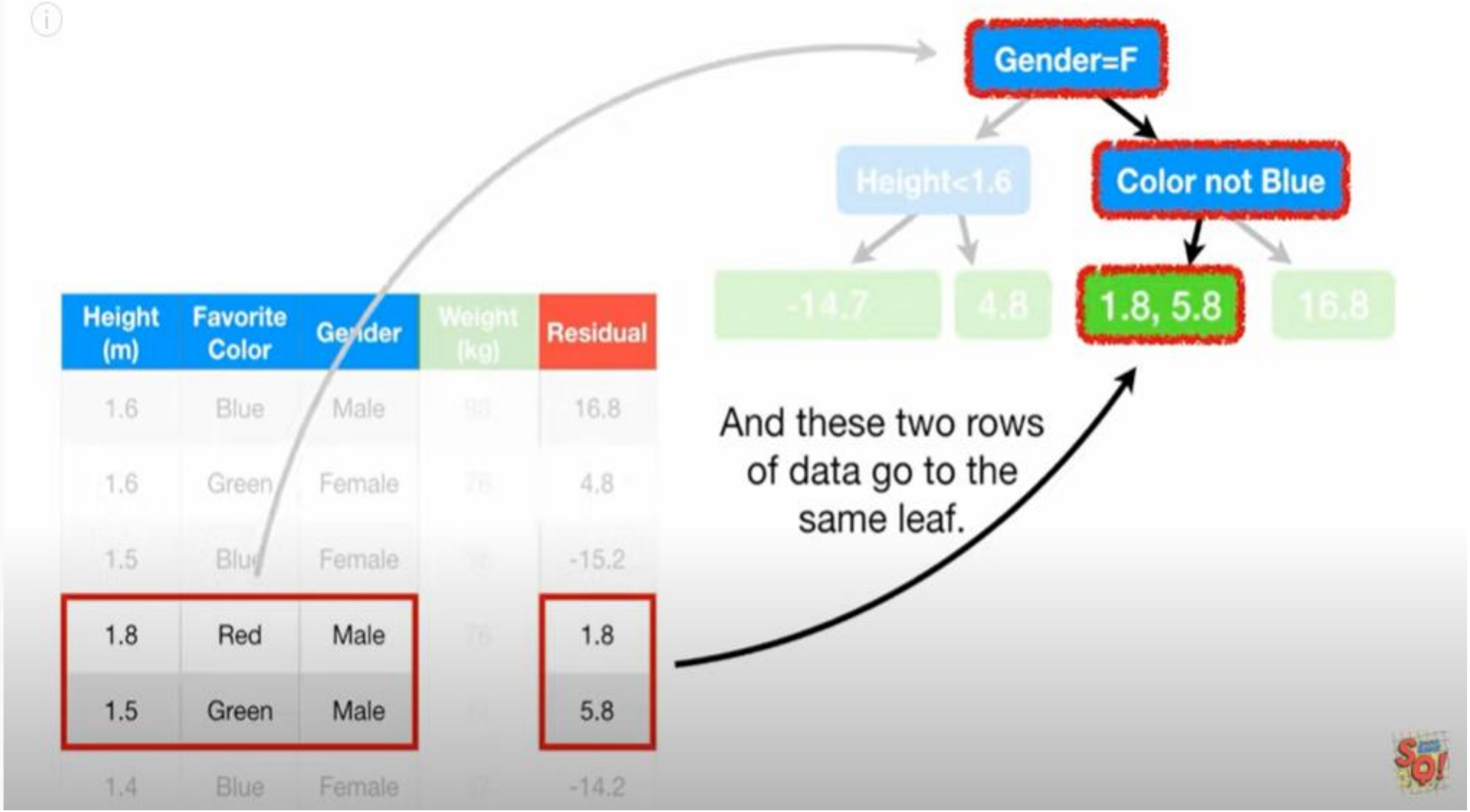

Gender=F
Height<1.6
Color not Blue
-14.2, -15.2
4.8
1.8, 5.8
16.8
Height (m)
Favorite Color
Gender
Residual
1.6
Blue
Male
16.8
1.6
Green
Female
4.8
1.5
Blue
Female
-15.2
1.8
Red
Male
1.8
1.5
Green
Male
5.8
1.4
Blue
Female
-14.2
As a result, these two rows of data go to the same leaf.

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	[illegible]	16.8
1.6	Green	Female	[illegible]	4.8
1.5	Blue	Female	[illegible]	-15.2
1.8	Red	Male	[illegible]	1.8
1.5	Green	Male	[illegible]	5.8
1.4	Blue	Female	[illegible]	-14.2



So we replace these residuals with their average.

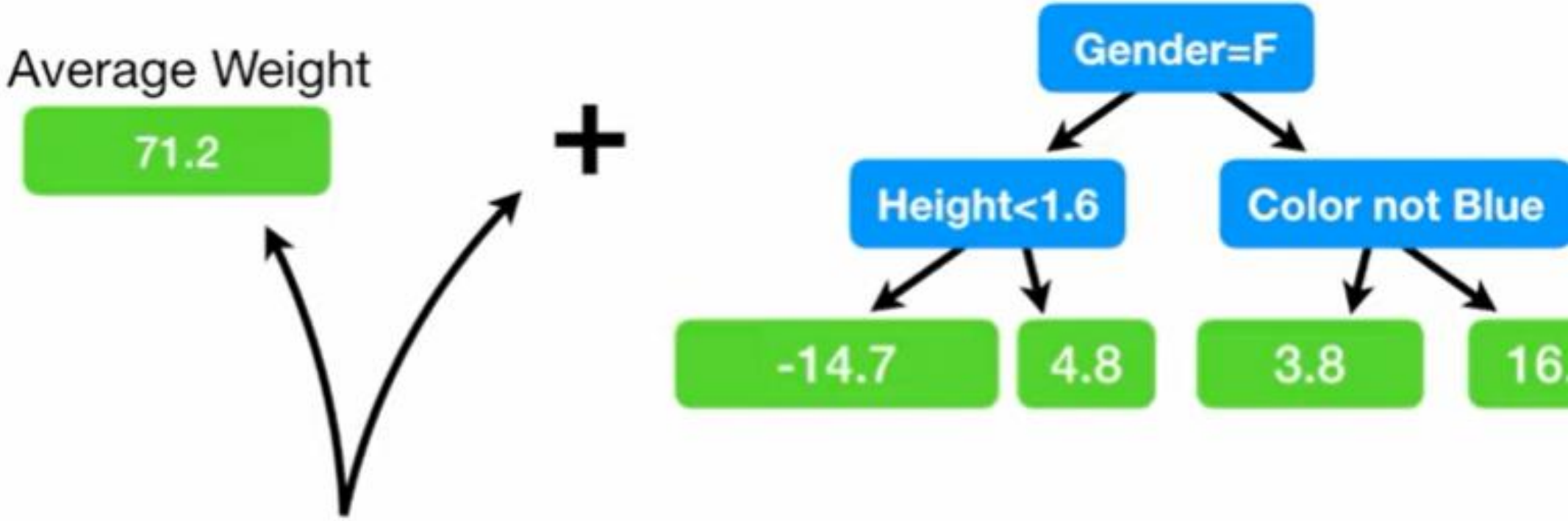
$$\frac{(-14.2 + -15.2)}{2} = -14.7$$


Gender=F
Height<1.6
Color not Blue
-14.7
4.8
1.8, 5.8
16.8
Height (m)
Favorite Color
Gender
Weight (kg)
Residual
1.6
Blue
Male
16.8
1.6
Green
Female
4.8
1.5
Blue
Female
-15.2
1.8
Red
Male
1.8
1.5
Green
Male
5.8
1.4
Blue
Female
-14.2
And these two rows of data go to the same leaf.

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	[illegible]	16.8
1.6	Green	Female	[illegible]	4.8
1.5	Blue	Female	[illegible]	-15.2
1.8	Red	Male	[illegible]	1.8
1.5	Green	Male	[illegible]	5.8
1.4	Blue	Female	[illegible]	-14.2

So we replace these residuals with their average.

$$\frac{(1.8 + 5.8)}{2} = 3.8$$

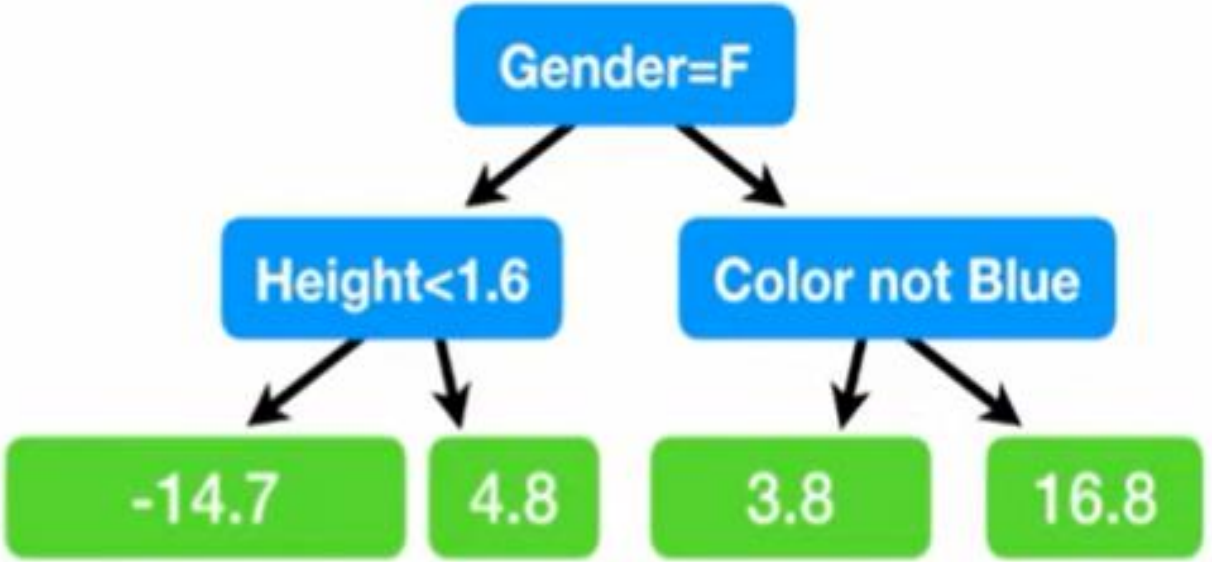


Now we can now combine the original leaf. ...with the new tree...

Average Weight

71.2

+



Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88

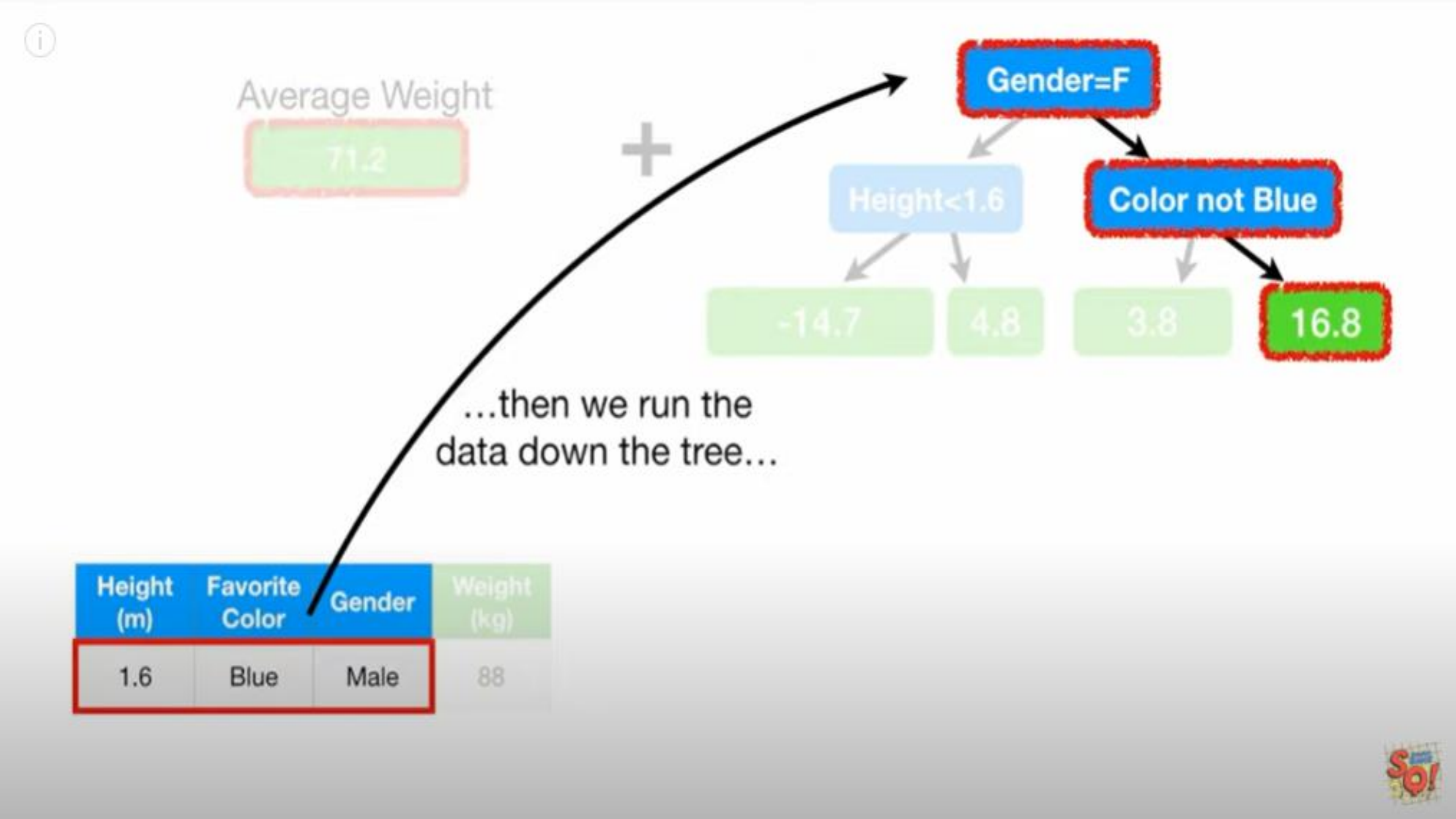
…to make a new **Prediction** of an individual's **Weight** from the **Training Data**.

Average Weight

71.2

We start with the initial **Prediction, 71.2…**

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88


Average Weight
71.2
+
Gender=F
Height<1.6
Color not Blue
-14.7
4.8
3.8
16.8
…then we run the data down the tree…
Height (m)
Favorite Color
Gender
Weight (kg)
1.6
Blue
Male
88

Predicted Weight = 71.2 + 16.8 = 88

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88

…which is the same as the **Observed Weight**.

Predicted Weight = 71.2 + 16.8 = 88

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88

In other words, we have low **Bias**, but probably very high **Variance**.

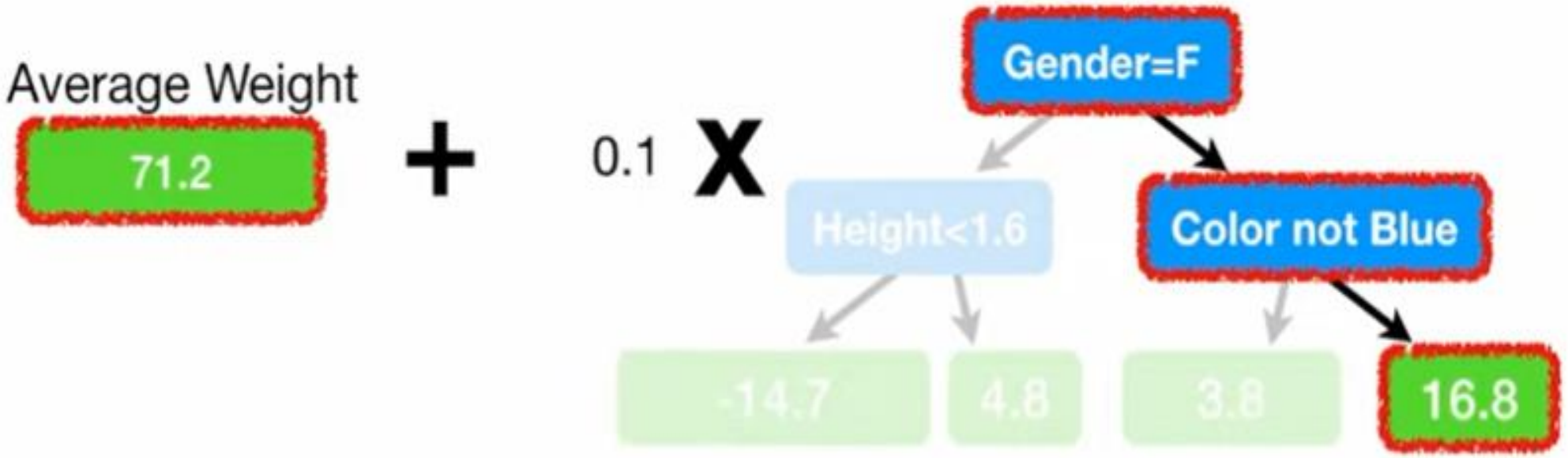
Average Weight

71.2

+ Learning Rate X

Gradient Boost deals with this problem by using a **Learning Rate** to scale the contribution from the new tree.

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88



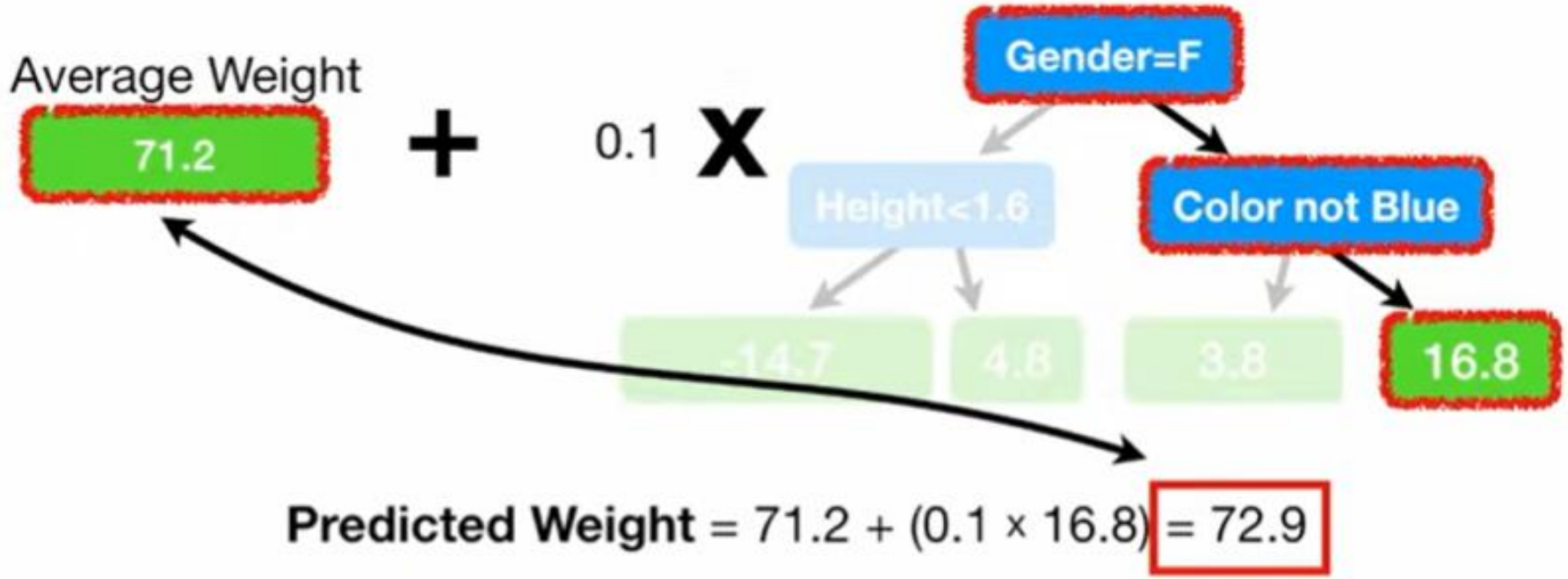
Now the **Predicted Weight** = 71.2 + (0.1 × 16.8) = 72.9

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88

Predicted Weight = 71.2 + (0.1 × 16.8) = 72.9

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88

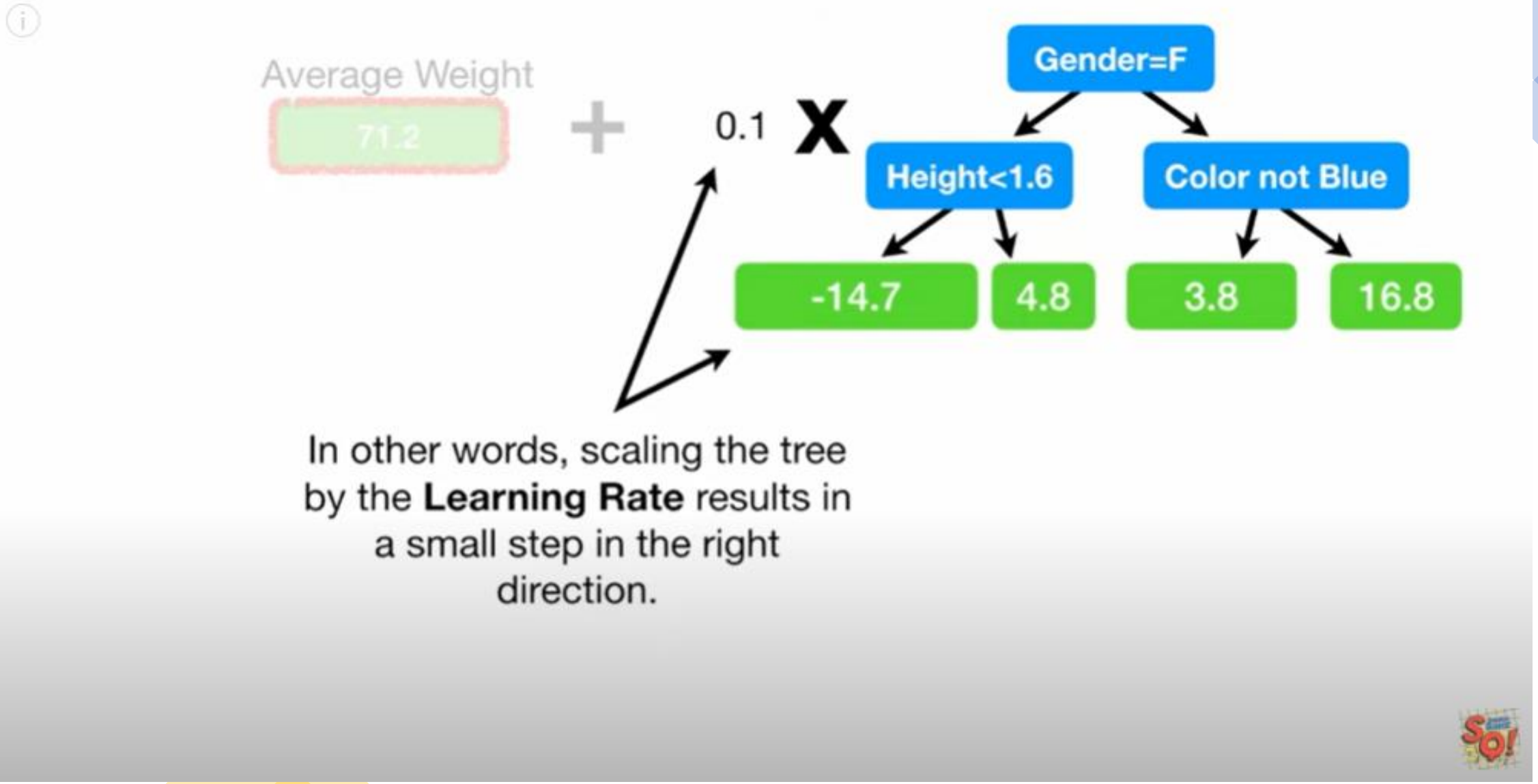
With the **Learning Rate** set to **0.1**, the new **Prediction** isn't as good as as it was before...

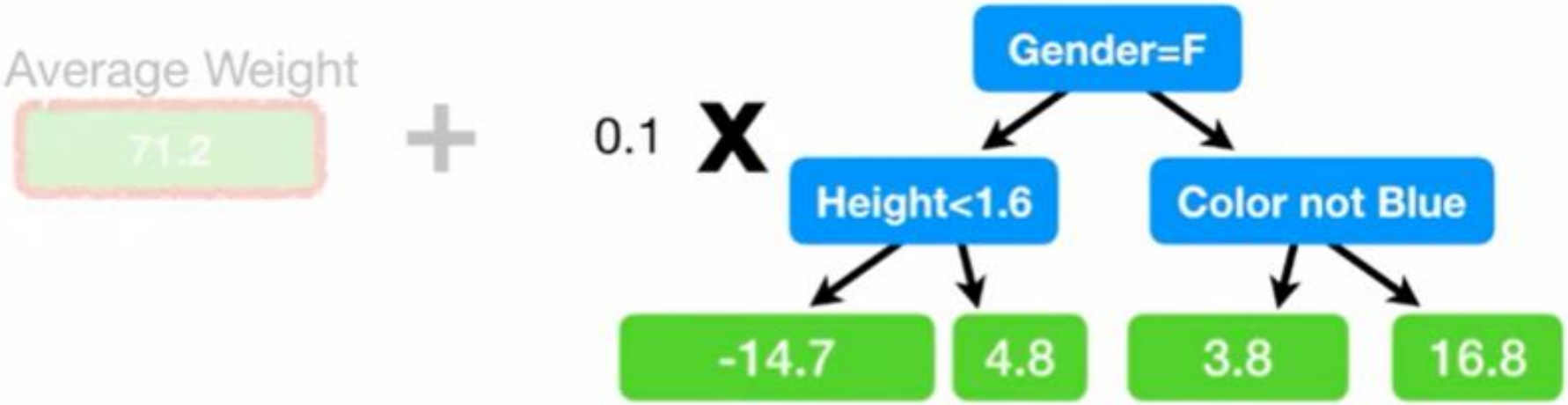


Predicted Weight = 71.2 + (0.1 × 16.8) = 72.9

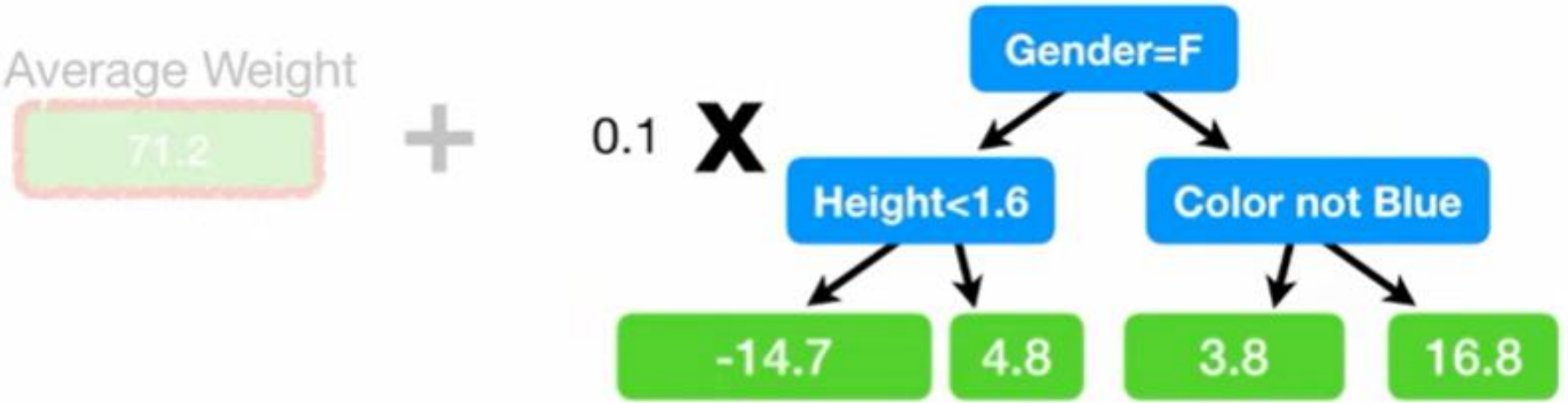
Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88

…but it's a little bit better than the **Prediction** made with just the original leaf, which predicted that all samples would weigh **71.2**.


Average Weight
71.2
+
0.1
X
Gender=F
Height<1.6
Color not Blue
-14.7
4.8
3.8
16.8
In other words, scaling the tree by the **Learning Rate** results in a small step in the right direction.



According to the dude that invented **Gradient Boost**, Jerome Friedman, empirical evidence shows that taking lots of small steps in the right direction results in better **Predictions** with a **Testing Dataset**, i.e. lower **Variance**.



So let's build another tree so we can take another small step in the right direction.

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	
1.6	Green	Female	76	
1.5	Blue	Female	56	
1.8	Red	Male	73	
1.5	Green	Male	77	
1.4	Blue	Female	57	

Just like before, we calculate the **Pseudo Residuals**, the difference between the **Observed Weights** and our latest **Predictions**.

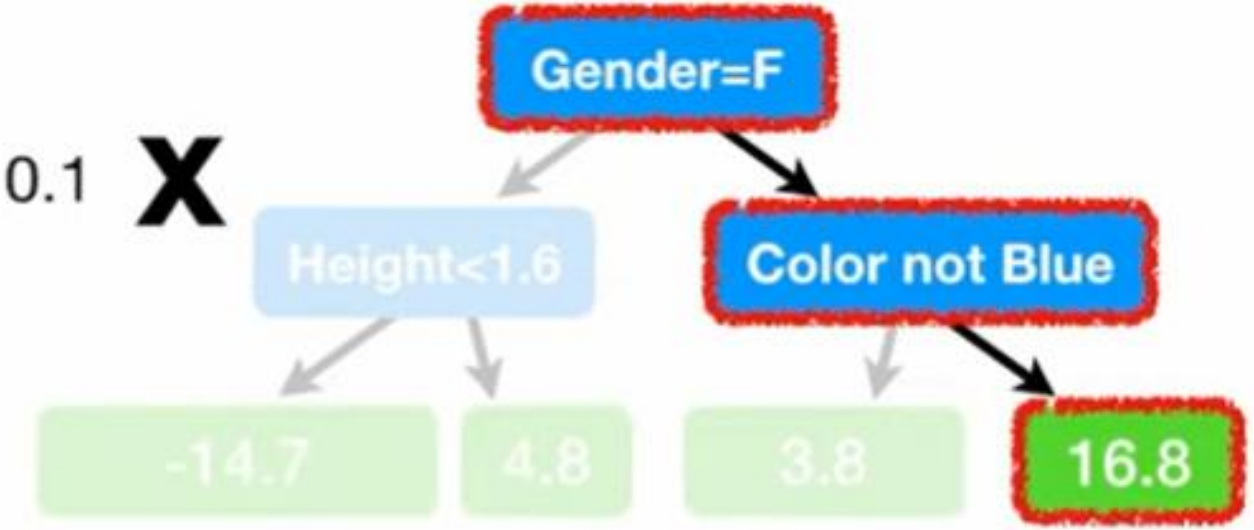
← **Residual** = (**Observed** - **Predicted**)

Average Weight

71.2

+

0.1 X



Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	
1.6	Green	Female	76	
1.5	Blue	Female	56	
1.8	Red	Male	73	
1.5	Green	Male	77	
1.4	Blue	Female	57	

Residual = (**88** - (71.2 + 0.1 × 16.8))

= 15.1

…and we get **15.1**…

Average Weight

71.2

+

0.1 X

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	15.1
1.6	Green	Female	76	
1.5	Blue	Female	56	
1.8	Red	Male	73	
1.5	Green	Male	77	
1.4	Blue	Female	57	

Residual = (**88** - (71.2 + 0.1 × 16.8))

= 15.1

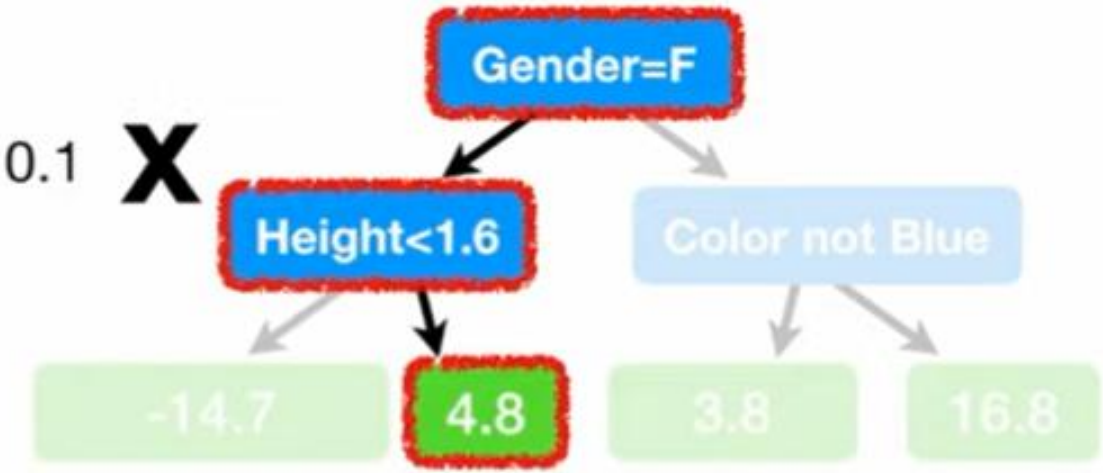
…and we save that in the column for **Pseudo Residuals**.

Average Weight

71.2

+

0.1 X



Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	15.1
1.6	Green	Female	76	4.3
1.5	Blue	Female	56	
1.8	Red	Male	73	
1.5	Green	Male	77	
1.4	Blue	Female	57	

Residual = (**76** - 71.2 + (0.1 × 4.8))

= 4.3

Then we repeat for the all of the other individuals in the **Training Dataset**.

So!

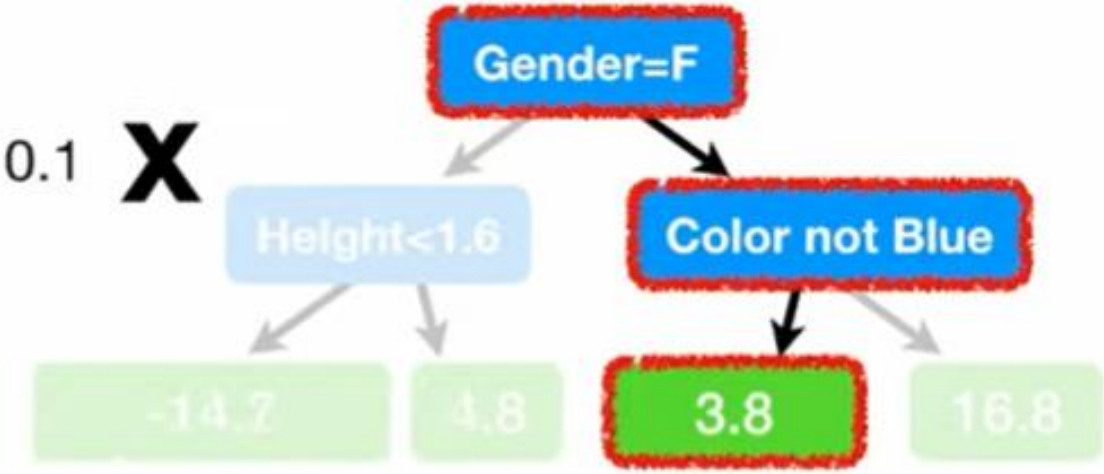
Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	15.1
1.6	Green	Female	76	4.3
1.5	Blue	Female	56	-13.7
1.8	Red	Male	73	
1.5	Green	Male	77	
1.4	Blue	Female	57	

Average Weight

71.2

+

0.1 X



Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	15.1
1.6	Green	Female	76	4.3
1.5	Blue	Female	56	-13.7
1.8	Red	Male	73	1.4
1.5	Green	Male	77	
1.4	Blue	Female	57	

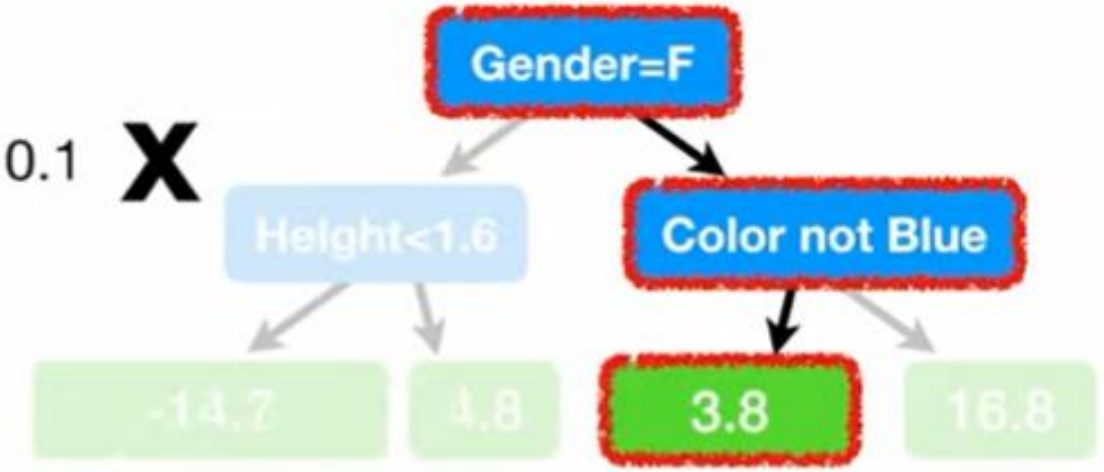
Residual = (Observed - Predicted)

Average Weight

71.2

+

0.1 X



Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	15.1
1.6	Green	Female	76	4.3
1.5	Blue	Female	56	-13.7
1.6	Red	Male	73	1.4
1.5	Green	Male	77	5.4
1.4	Blue	Female	57	

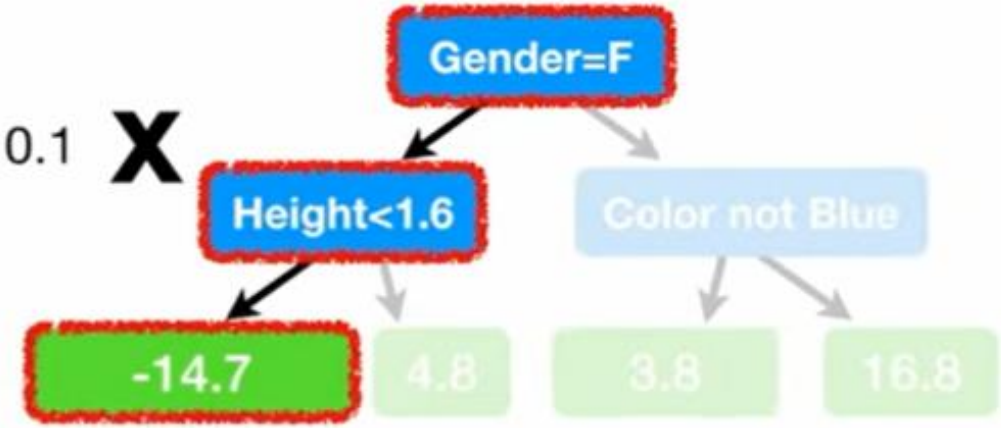
Residual = (Observed - Predicted)

Average Weight

71.2

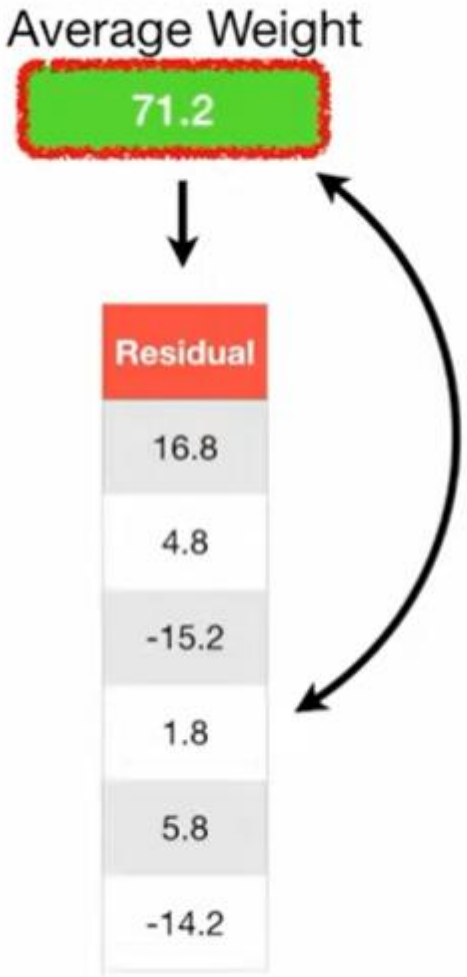
+

0.1 X

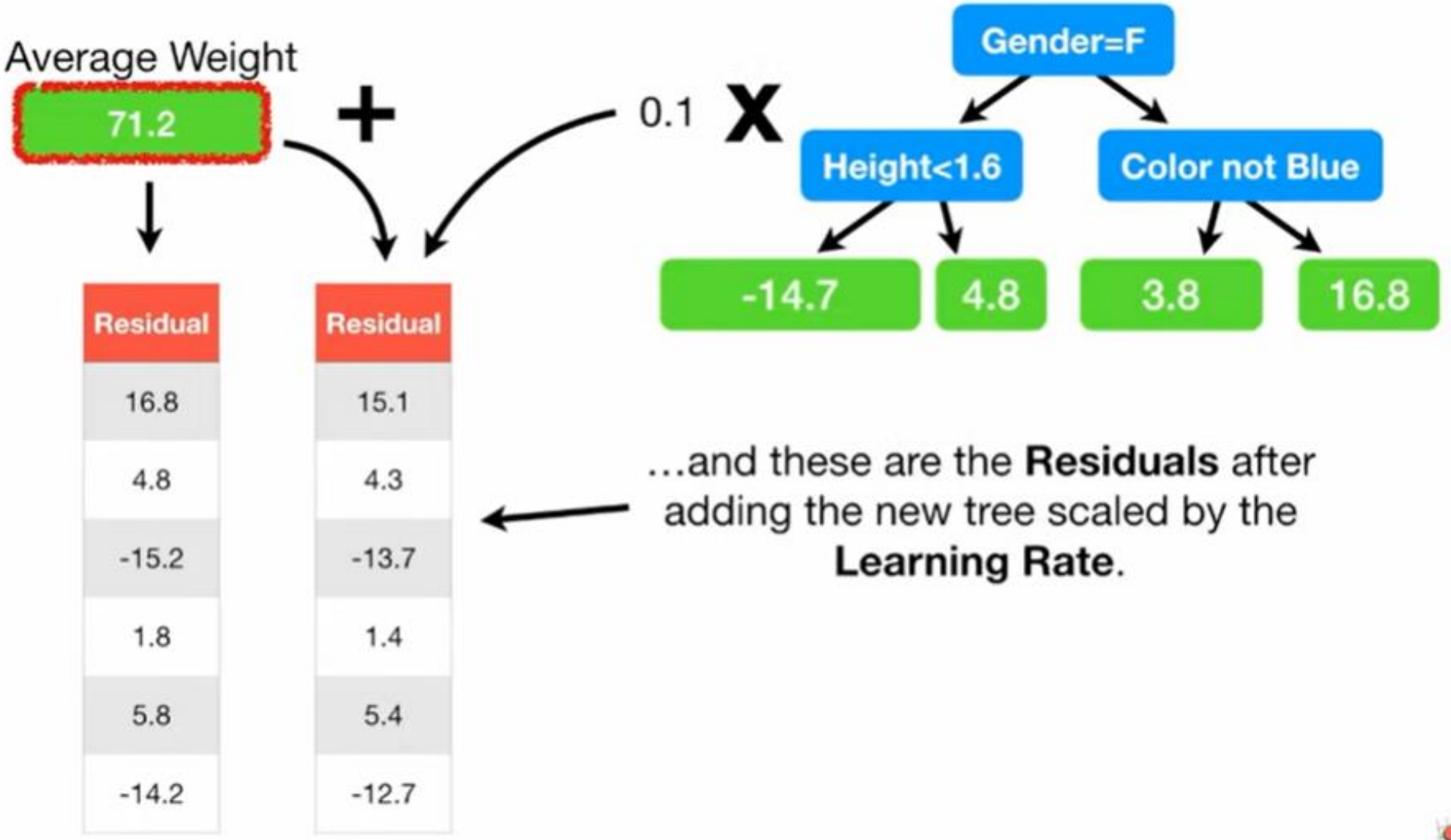


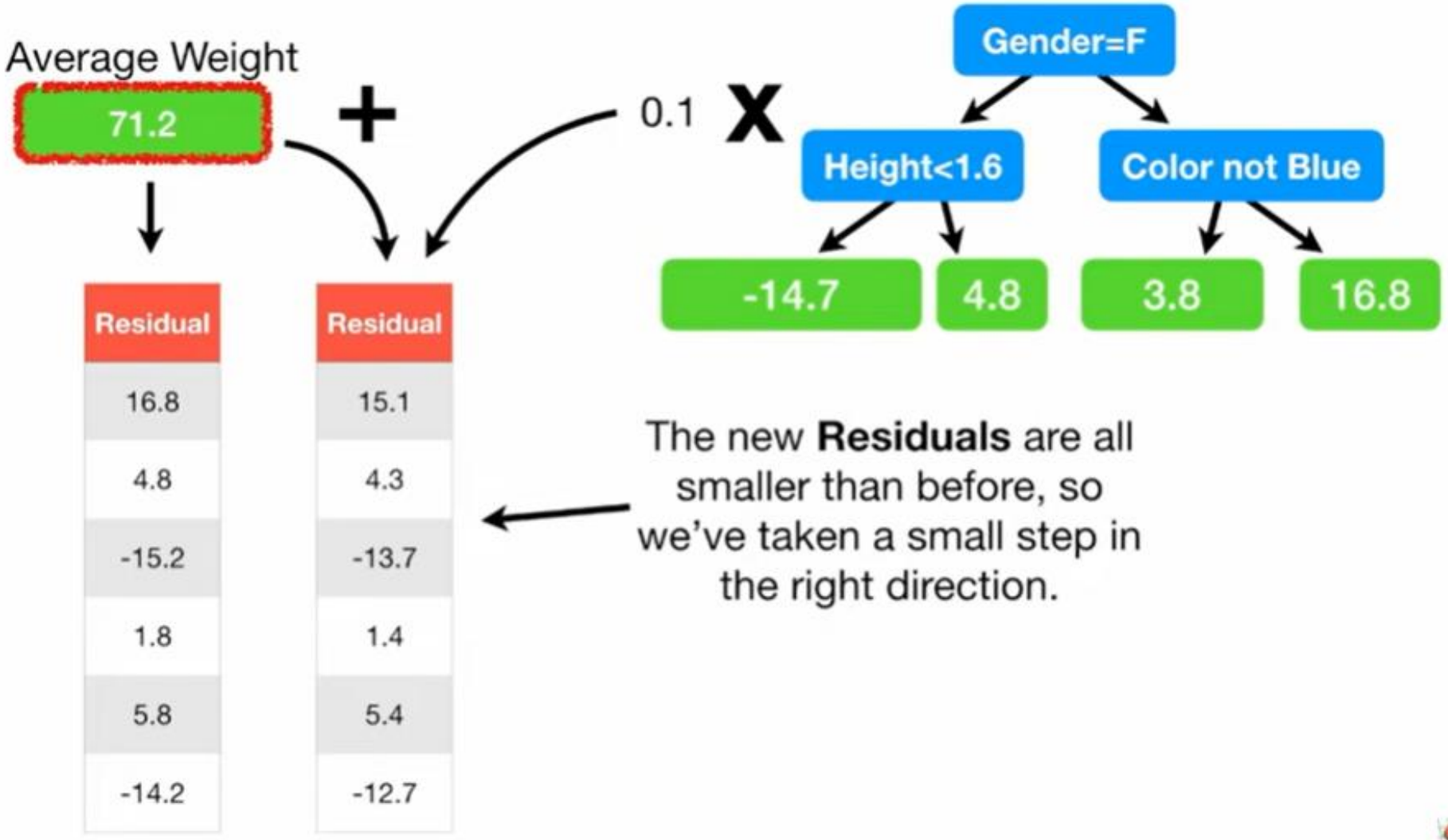
Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	15.1
1.6	Green	Female	76	4.3
1.5	Blue	Female	56	-13.7
1.8	Red	Male	73	1.4
1.5	Green	Male	77	5.4
1.4	Blue	Female	57	-12.7

Residual = (Observed - Predicted)

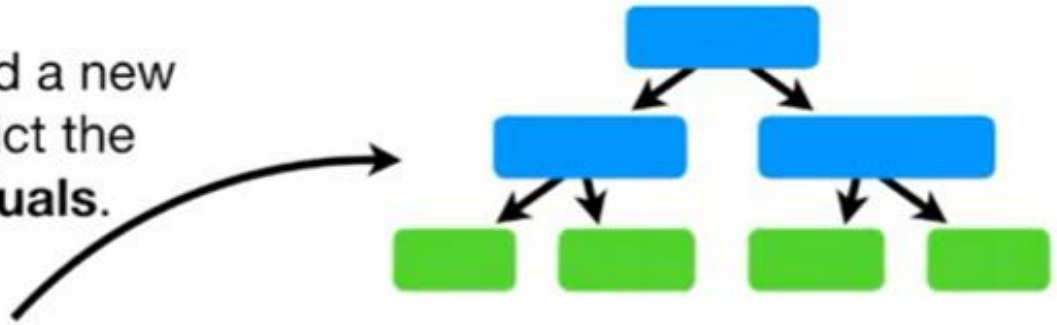


NOTE: These are the original **Residuals**, from when our **Prediction** was simply the average overall **Weight**.


Average Weight
71.2
+
0.1
X
Gender=F
Height<1.6
Color not Blue
-14.7
4.8
3.8
16.8
Residual
16.8
4.8
-15.2
1.8
5.8
-14.2
Residual
15.1
4.3
-13.7
1.4
5.4
-12.7
...and these are the **Residuals** after adding the new tree scaled by the **Learning Rate**.

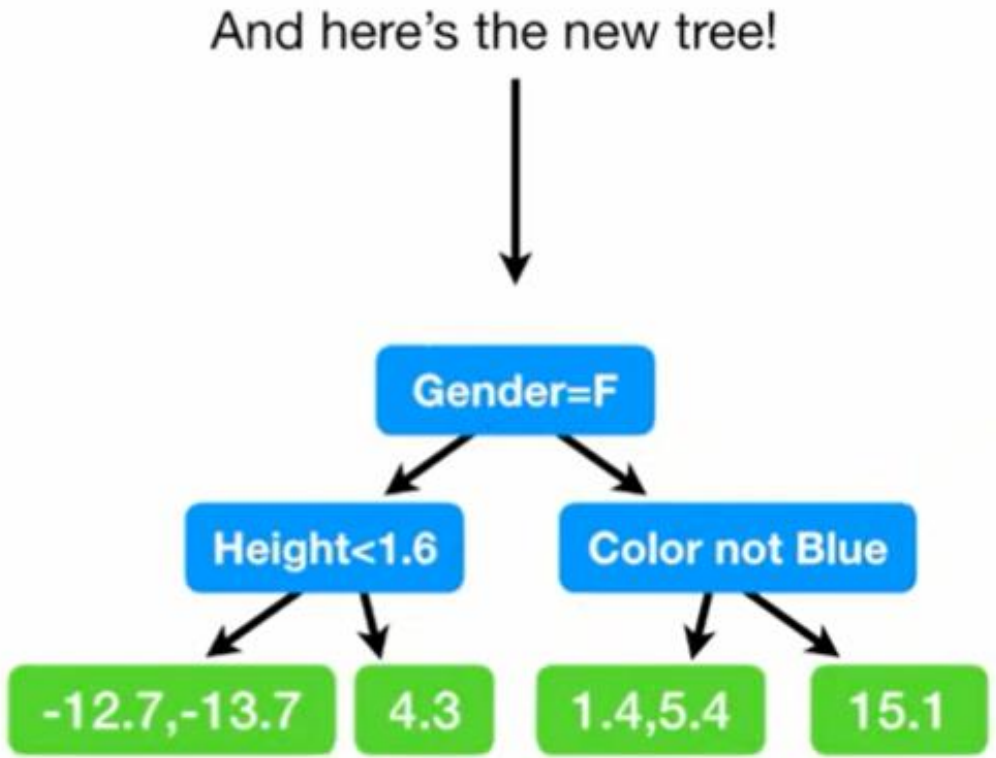

Average Weight
71.2
+
0.1
X
Gender=F
Height<1.6
Color not Blue
-14.7
4.8
3.8
16.8
Residual
16.8
4.8
-15.2
1.8
5.8
-14.2
Residual
15.1
4.3
-13.7
1.4
5.4
-12.7
The new **Residuals** are all smaller than before, so we've taken a small step in the right direction.

Now let's build a new tree to predict the new **Residuals**.



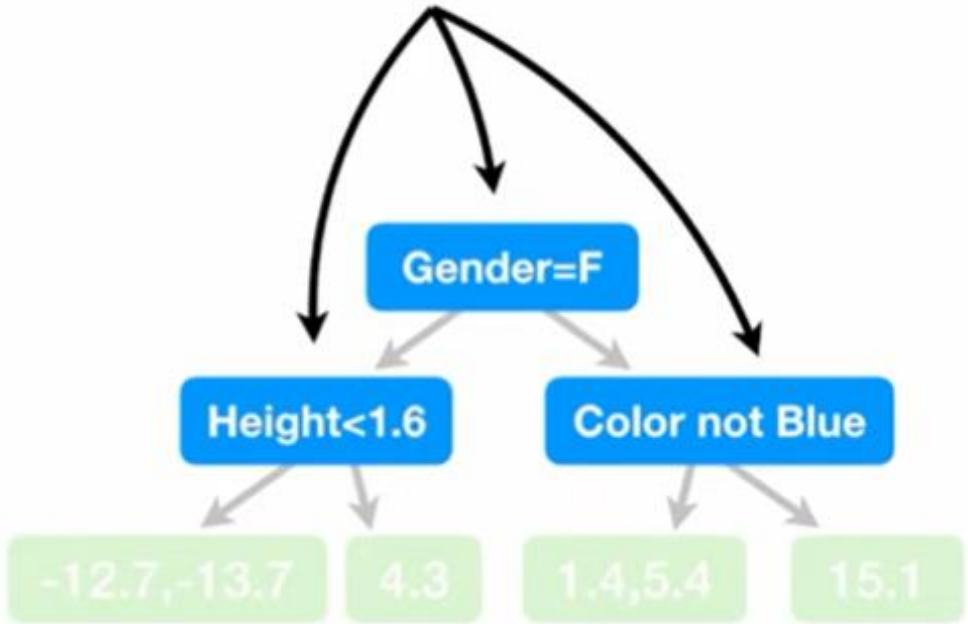
Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	15.1
1.6	Green	Female	76	4.3
1.5	Blue	Female	56	-13.7
1.8	Red	Male	73	1.4
1.5	Green	Male	77	5.4
1.4	Blue	Female	57	-12.7

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male		15.1
1.6	Green	Female		4.3
1.5	Blue	Female		-13.7
1.8	Red	Male		1.4
1.5	Green	Male		5.4
1.4	Blue	Female		-12.7

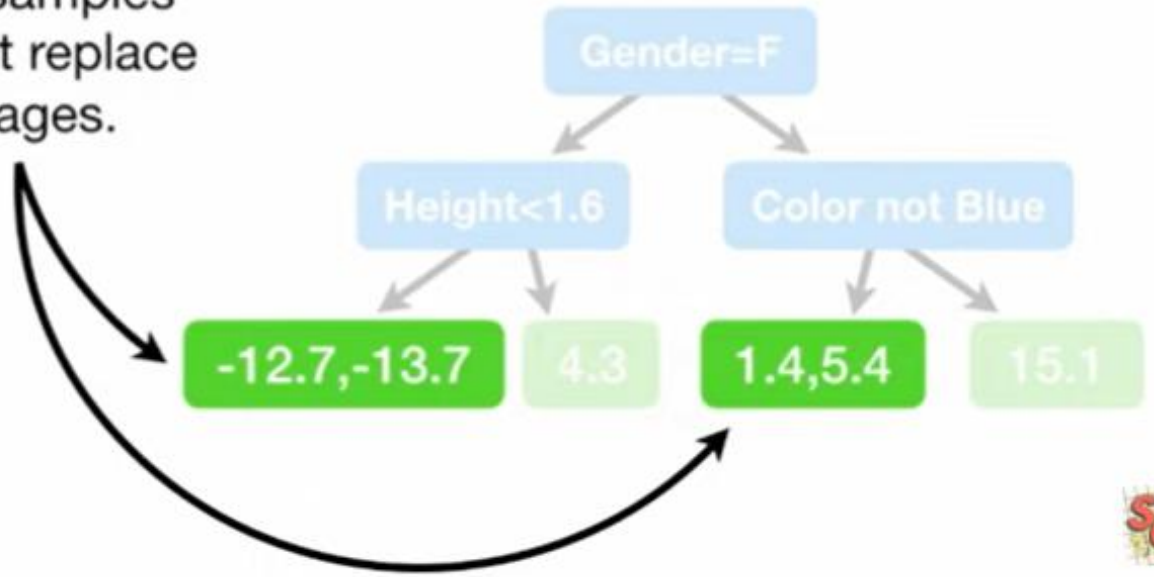


Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	15.1
1.6	Green	Female	76	4.3
1.5	Blue	Female	56	-13.7
1.8	Red	Male	73	1.4
1.5	Green	Male	77	5.4
1.4	Blue	Female	57	-12.7

NOTE: In this simple example the branches are the same as before. However, in practice, the trees can be different each time.



Just like before, since multiple samples ended up in these leaves, we just replace the **Residuals** with their averages.

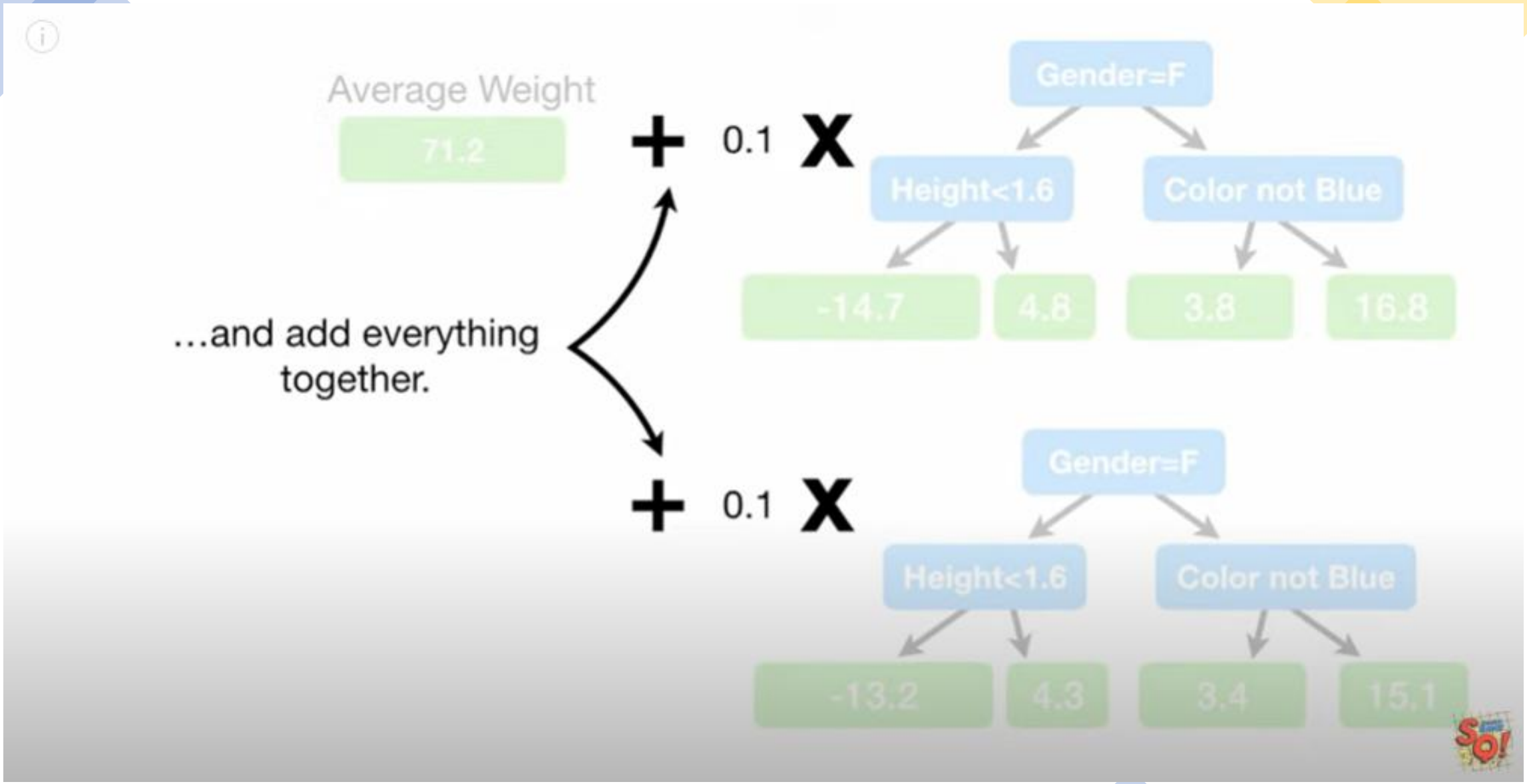


Average Weight

71.2

Now we combine the new **Tree** with the previous **Tree** and the initial **Leaf.**


Average Weight
71.2
0.1
X
Gender=F
Height<1.6
Color not Blue
-14.7
4.8
3.8
16.8
NOTE: We scale all of the Trees by the Learning Rate, which we set to 0.1…
0.1
X
Gender=F
Height<1.6
Color not Blue
-13.2
4.3
3.4
15.1


Average Weight
71.2
+
0.1
X
Gender=F
Height<1.6
Color not Blue
-14.7
4.8
3.8
16.8
…and add everything together.
+
0.1
X
Gender=F
Height<1.6
Color not Blue
-13.2
4.3
3.4
15.1


Average Weight
71.2
+ 0.1 X
Gender=F
Height<1.6
Color not Blue
-14.7
4.8
3.8
16.8
Now we're ready to make a new **Prediction** from the **Training Data**.
Height (m)
Favorite Color
Gender
Weight (kg)
1.6
Blue
Male
88
+ 0.1 X
Gender=F
Height<1.6
Color not Blue
-13.2
4.3
3.4
15.1
So!

Average Weight

71.2

+ 0.1 X

That gives us…

$$71.2 + (0.1 \times 16.8) + (0.1 \times 15.1)$$

+ 0.1 X

Average Weight

71.2

+ 0.1 X

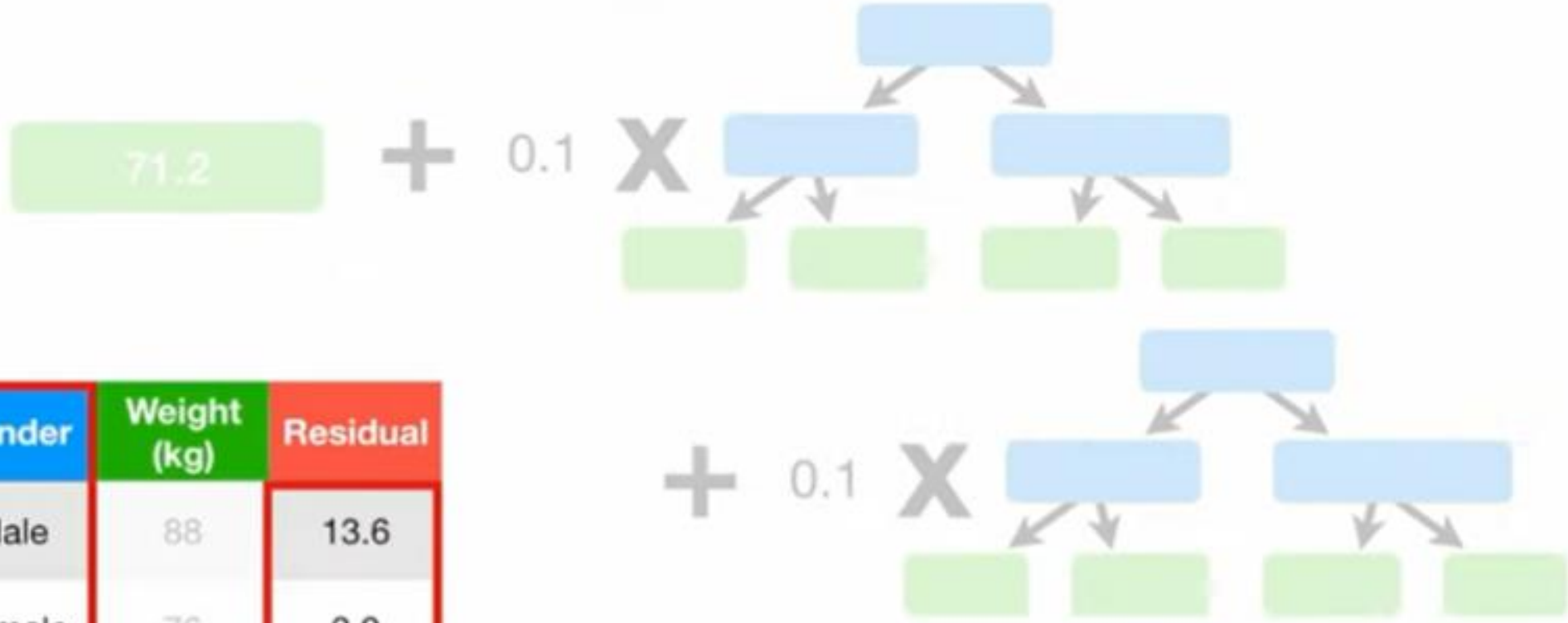
Which is another small step closer to the **Observed Weight**.

$71.2 + (0.1 \times 16.8) + (0.1 \times 15.1)$

$= 74.4$

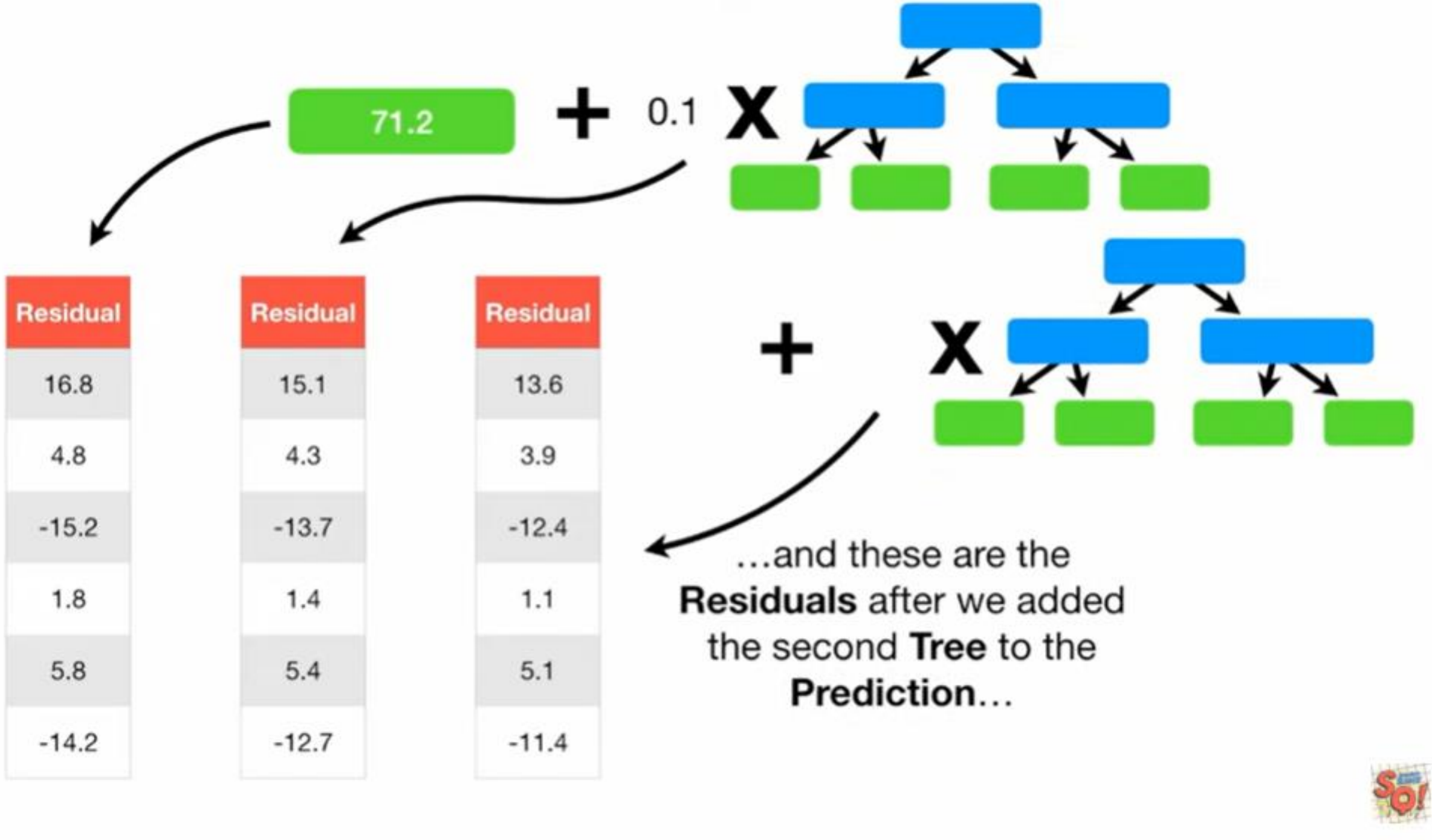
Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88

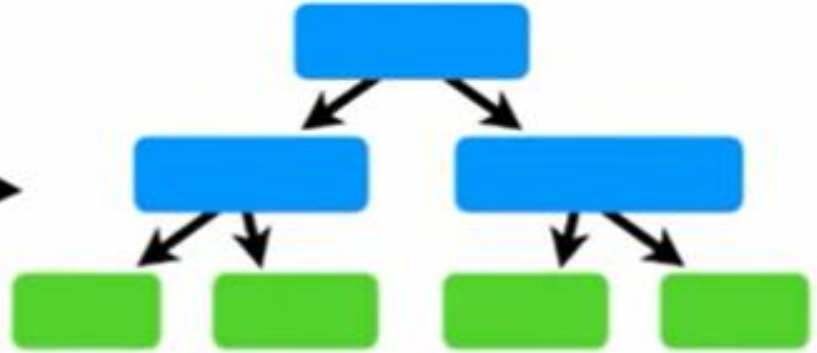
+ 0.1 X



Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	13.6
1.6	Green	Female	76	3.9
1.5	Blue	Female	56	-12.4
1.8	Red	Male	73	1.1
1.5	Green	Male	77	5.1
1.4	Blue	Female	57	-11.4

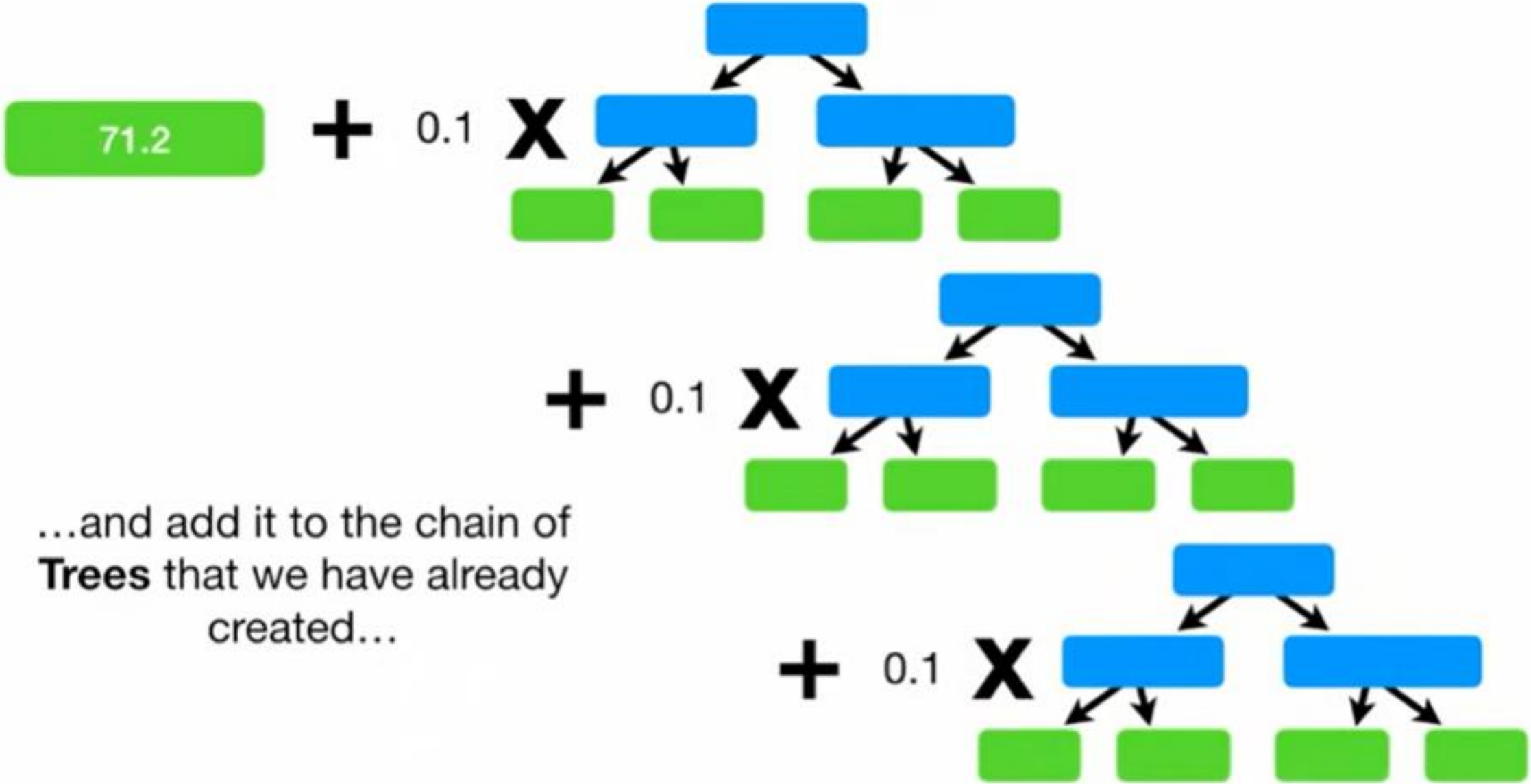
…to calculate new **Residuals.**


71.2
+
0.1
X
+
X
Residual
16.8
4.8
-15.2
1.8
5.8
-14.2
Residual
15.1
4.3
-13.7
1.4
5.4
-12.7
Residual
13.6
3.9
-12.4
1.1
5.1
-11.4
…and these are the **Residuals** after we added the second **Tree** to the **Prediction**…

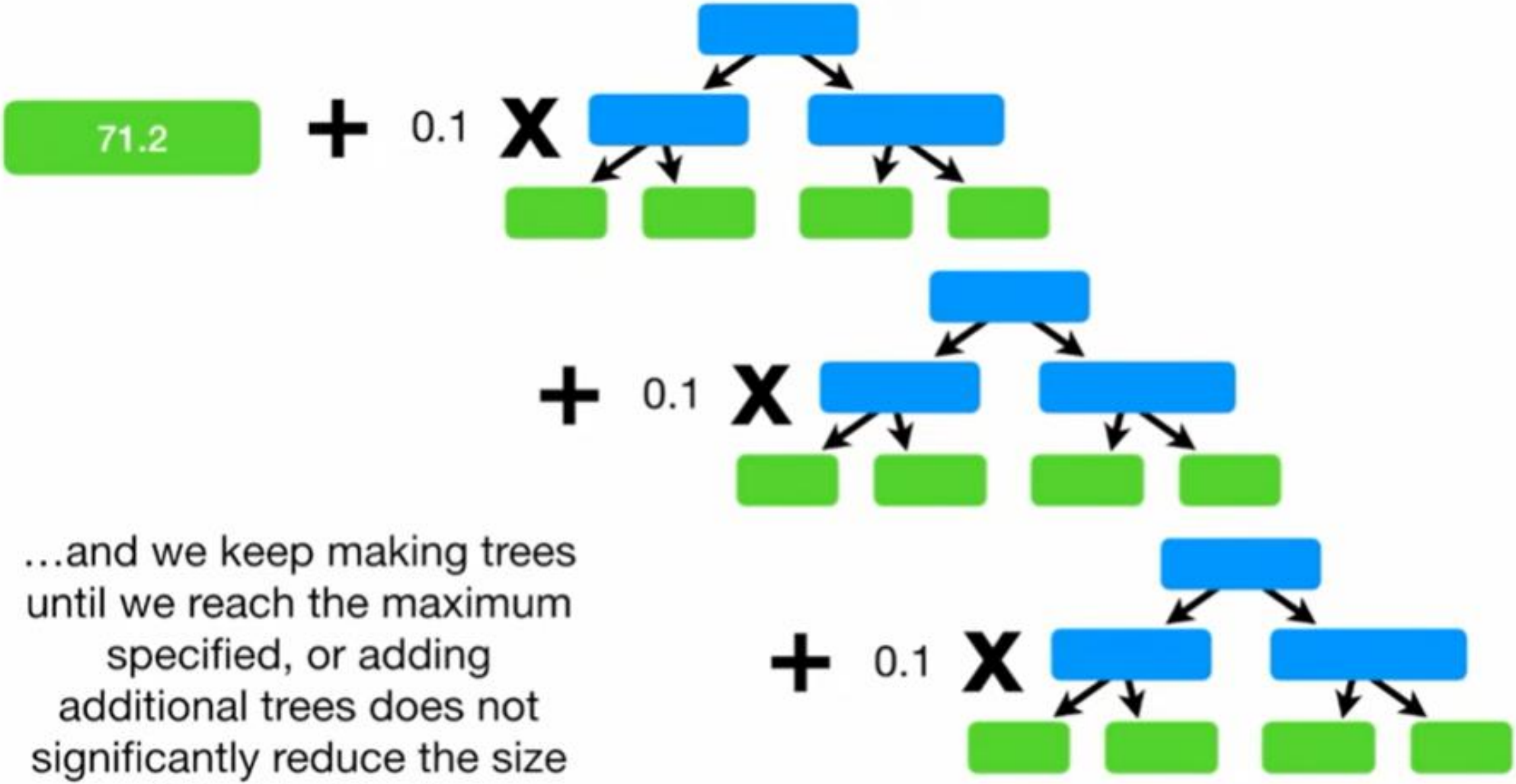


Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	13.6
1.6	Green	Female	76	3.9
1.5	Blue	Female	56	-12.4
1.8	Red	Male	73	1.1
1.5	Green	Male	77	5.1
1.4	Blue	Female	57	-11.4

Now we build a another tree to predict the new **Residuals**…

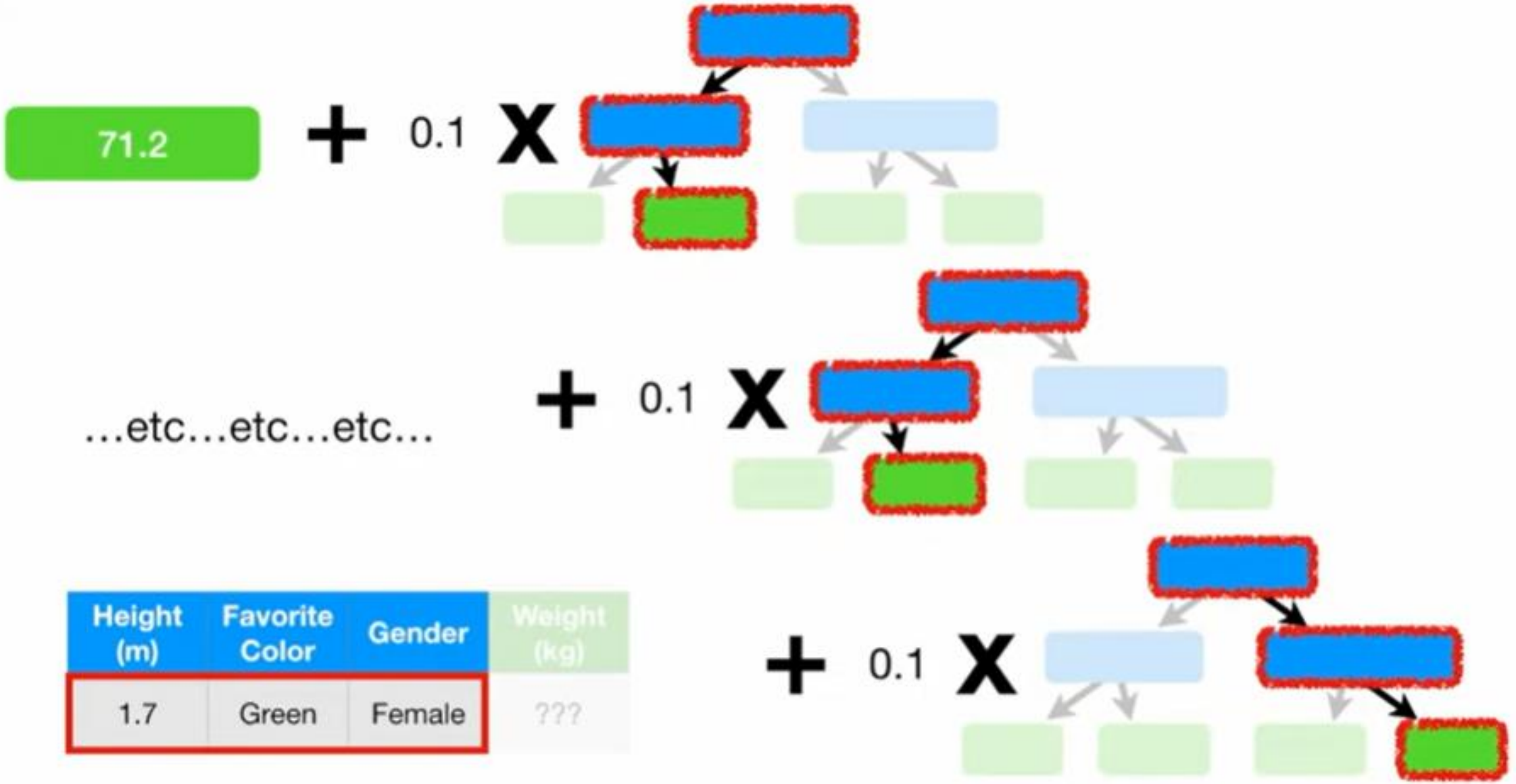


…and add it to the chain of **Trees** that we have already created…

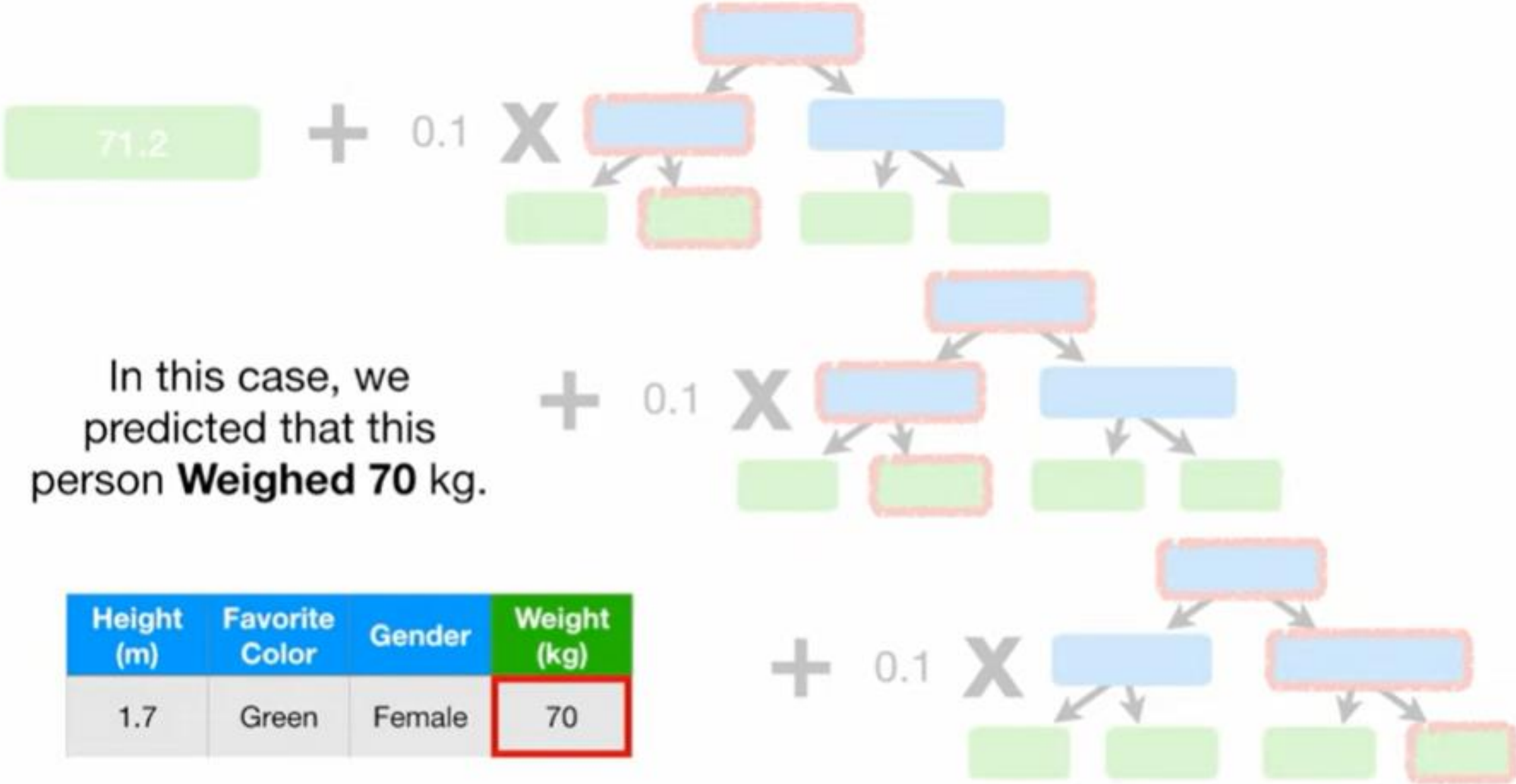


…and we keep making trees until we reach the maximum specified, or adding additional trees does not significantly reduce the size of the **Residuals**.

…etc…etc…etc…

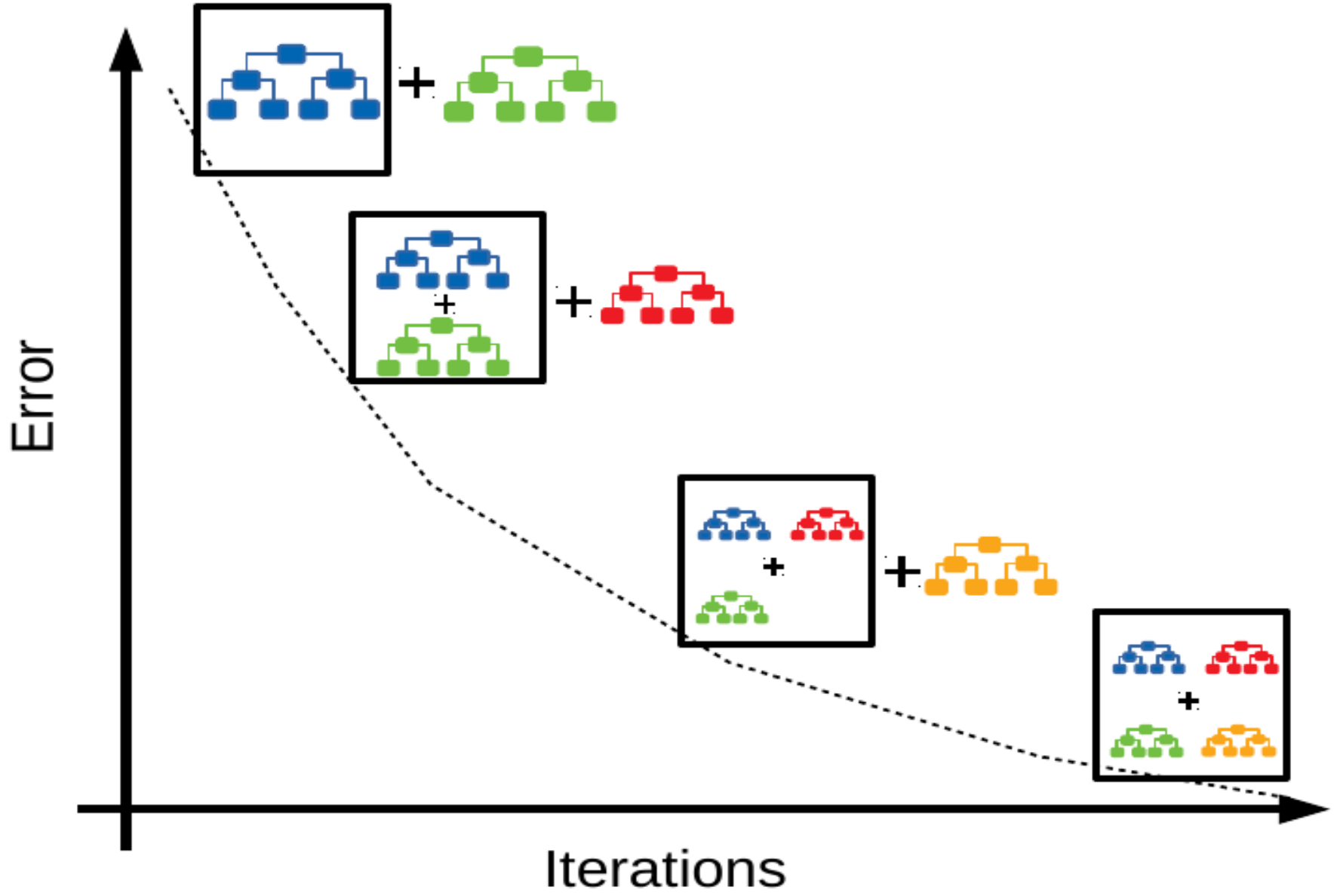


Height (m)	Favorite Color	Gender	Weight (kg)
1.7	Green	Female	???



In this case, we predicted that this person **Weighed 70** kg.

Height (m)	Favorite Color	Gender	Weight (kg)
1.7	Green	Female	70


Error
Iterations