…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

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 first thing we do is calculate the average **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

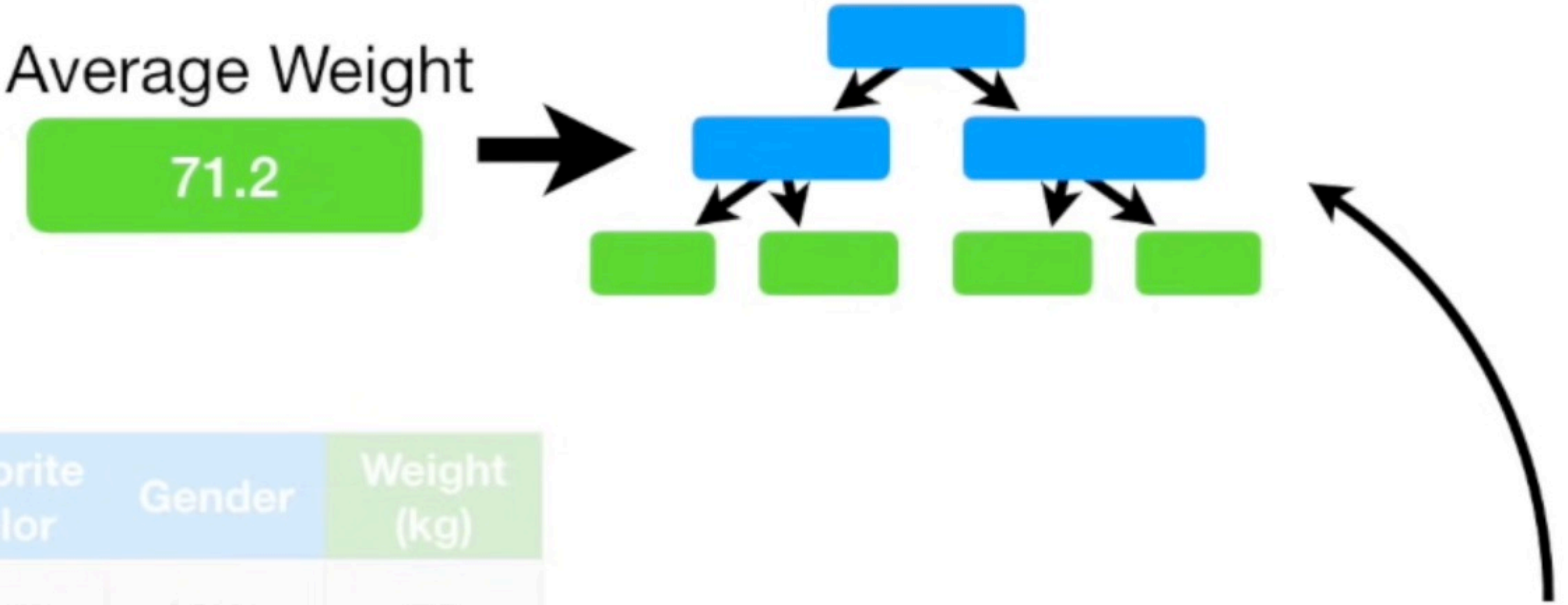
This is the first attempt at predicting everyone's weight.

Average Weight

71.2

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

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

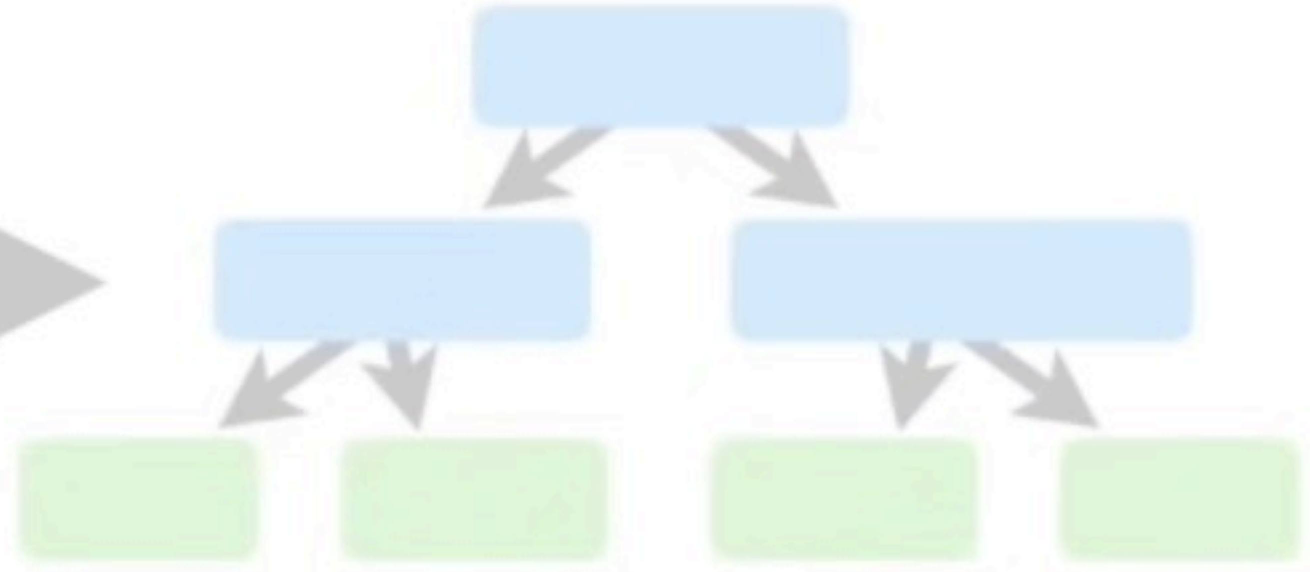


Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

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	$\boxed{16.8}$
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.

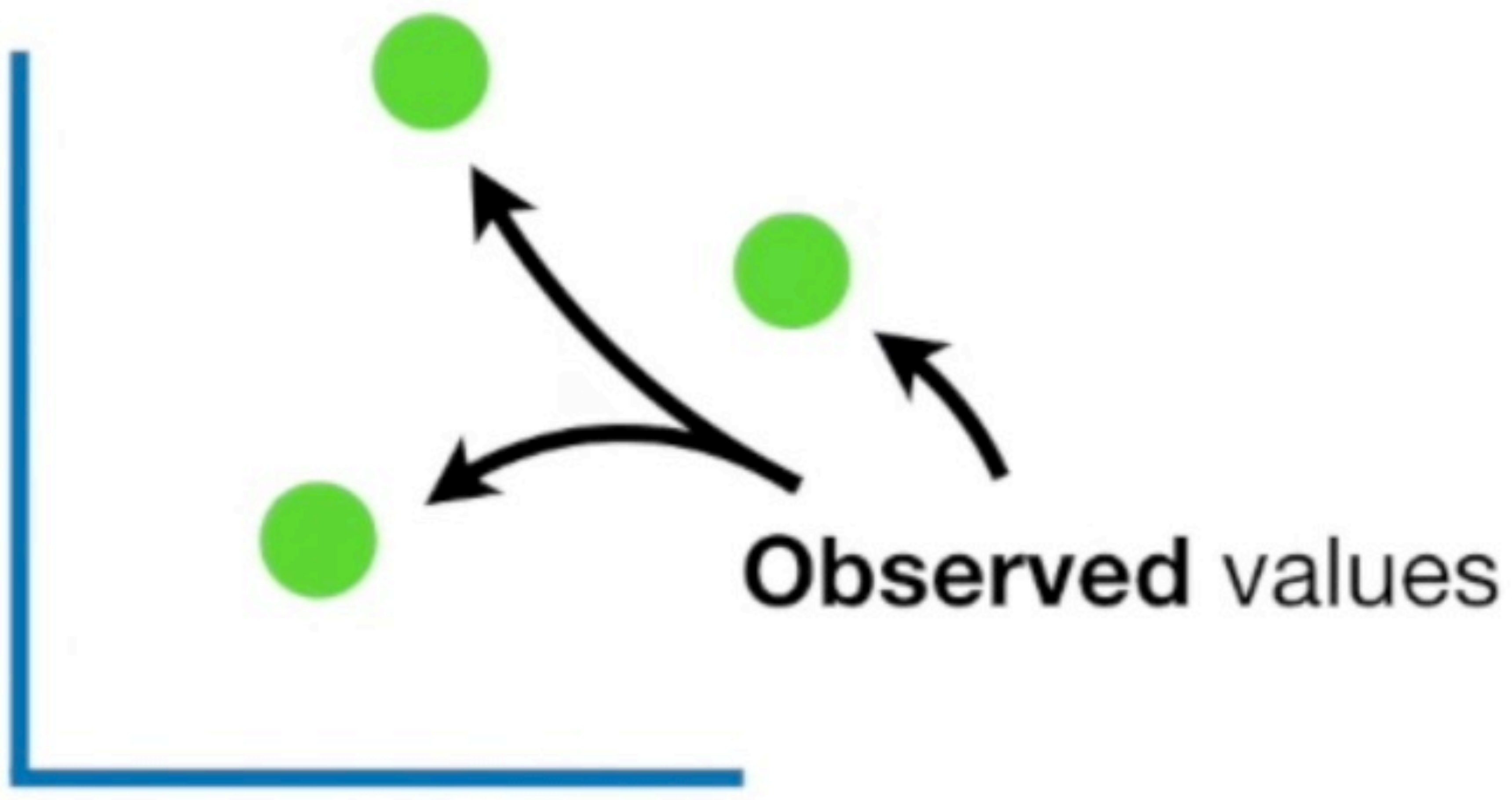
$$(\mathbf{88} - \mathbf{71.2}) = \boxed{16.8}$$

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	

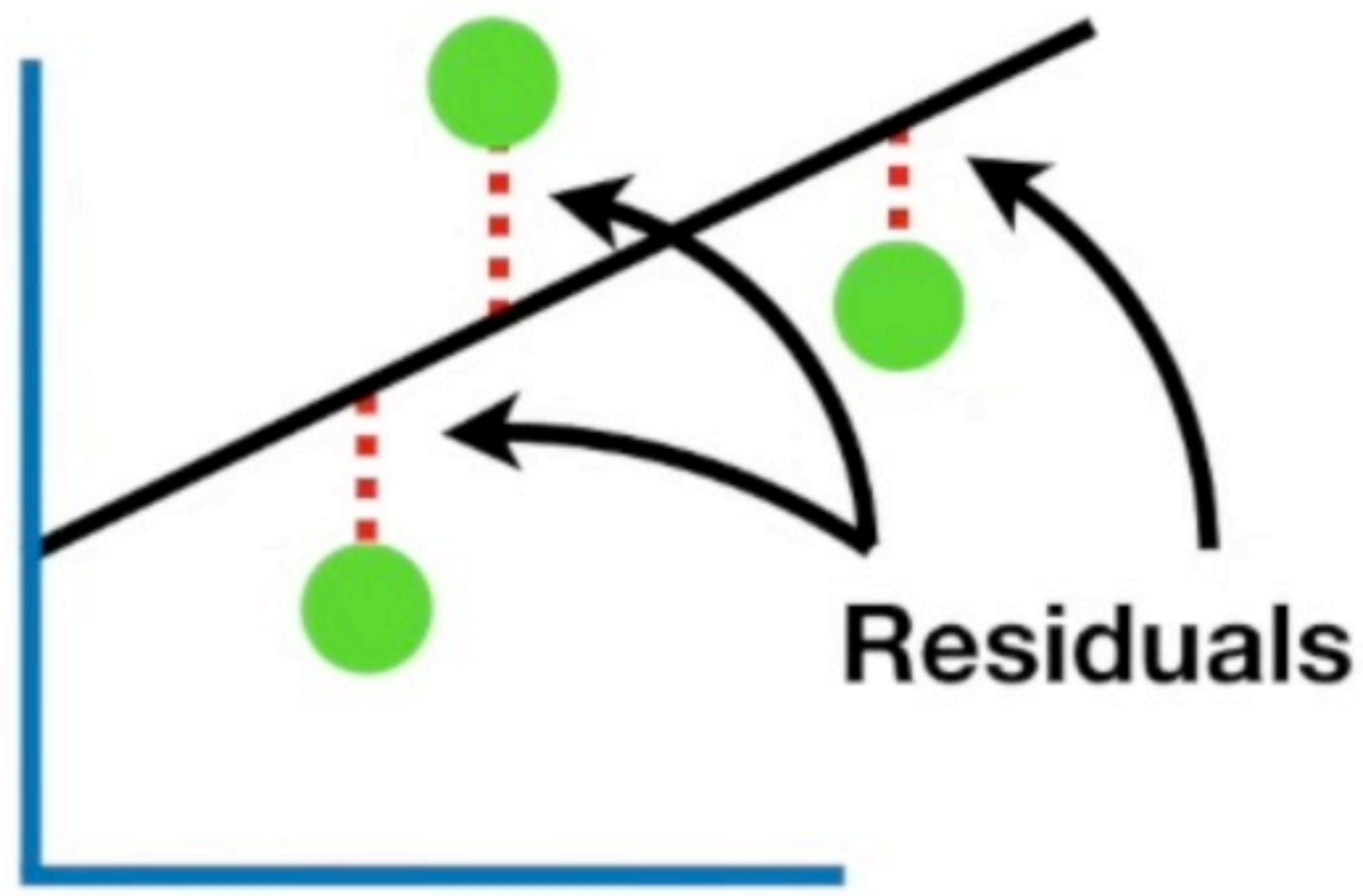


Average Weight

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	



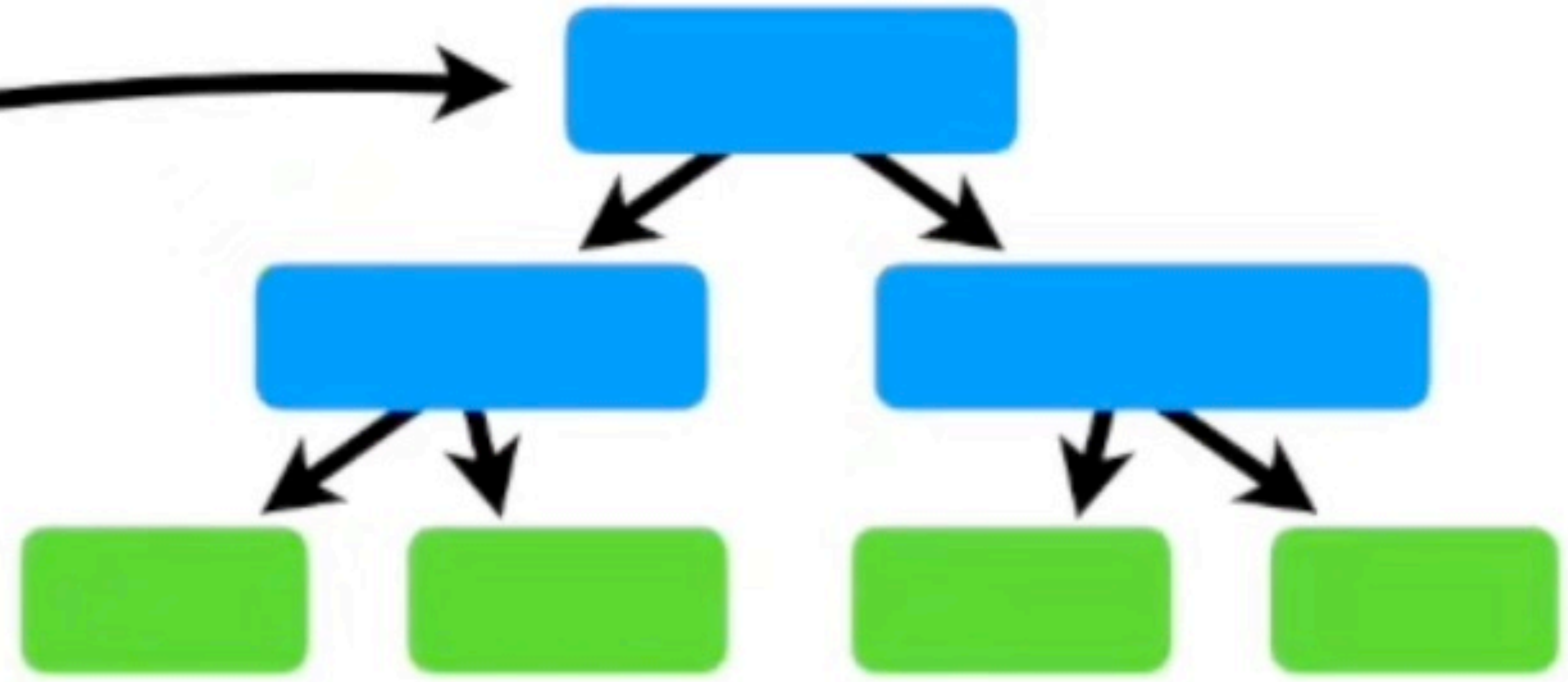
Average Weight

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	

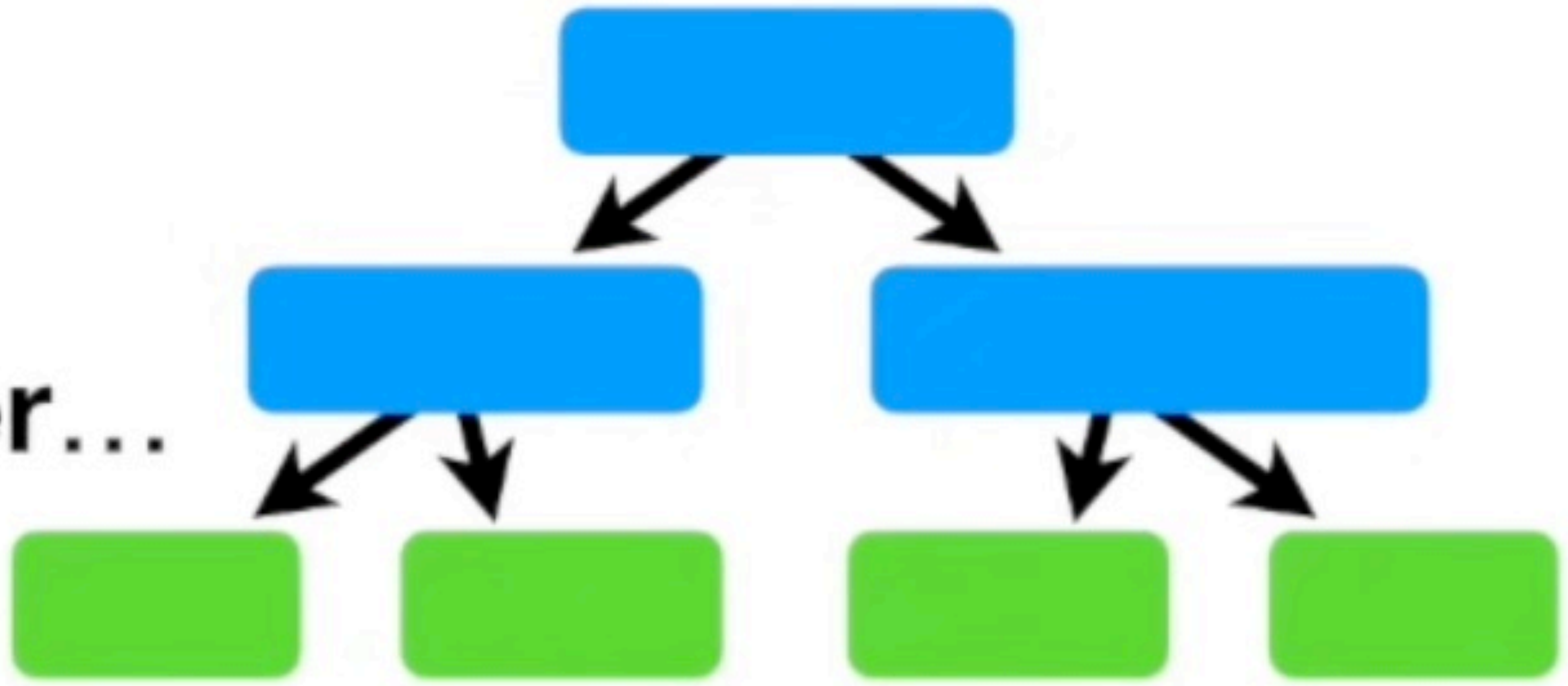
The “**Pseudo**” part of **Pseudo Residual** is a reminder that we are doing **Gradient Boost**, not **Linear Regression**, and is something I’ll talk more about in **Part 2** of this series when we go through the math.

Now we will build a **Tree**

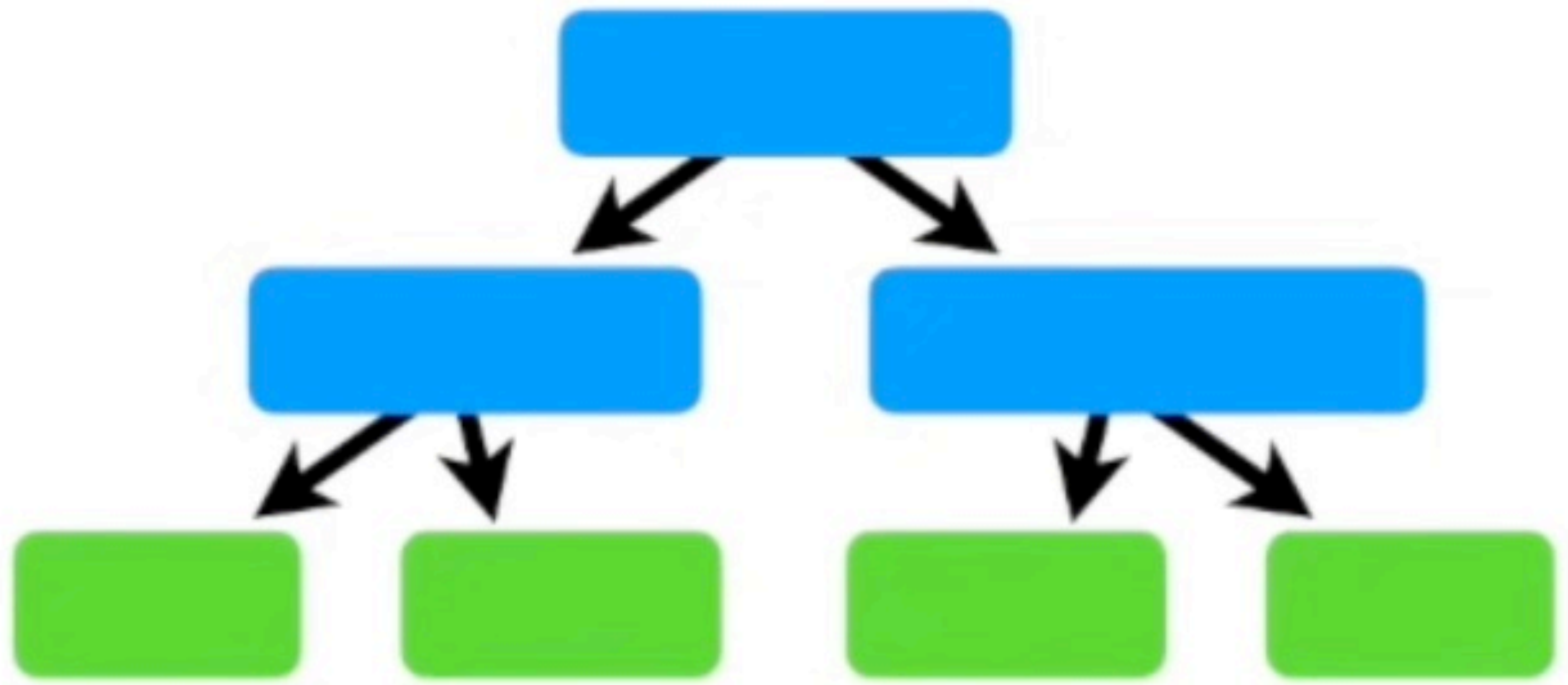


Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
1.5	Blue	Female	56	-15.2
1.8	Red	Male	73	1.8
1.5	Green	Male	77	5.8
1.4	Blue	Female	57	-14.2

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



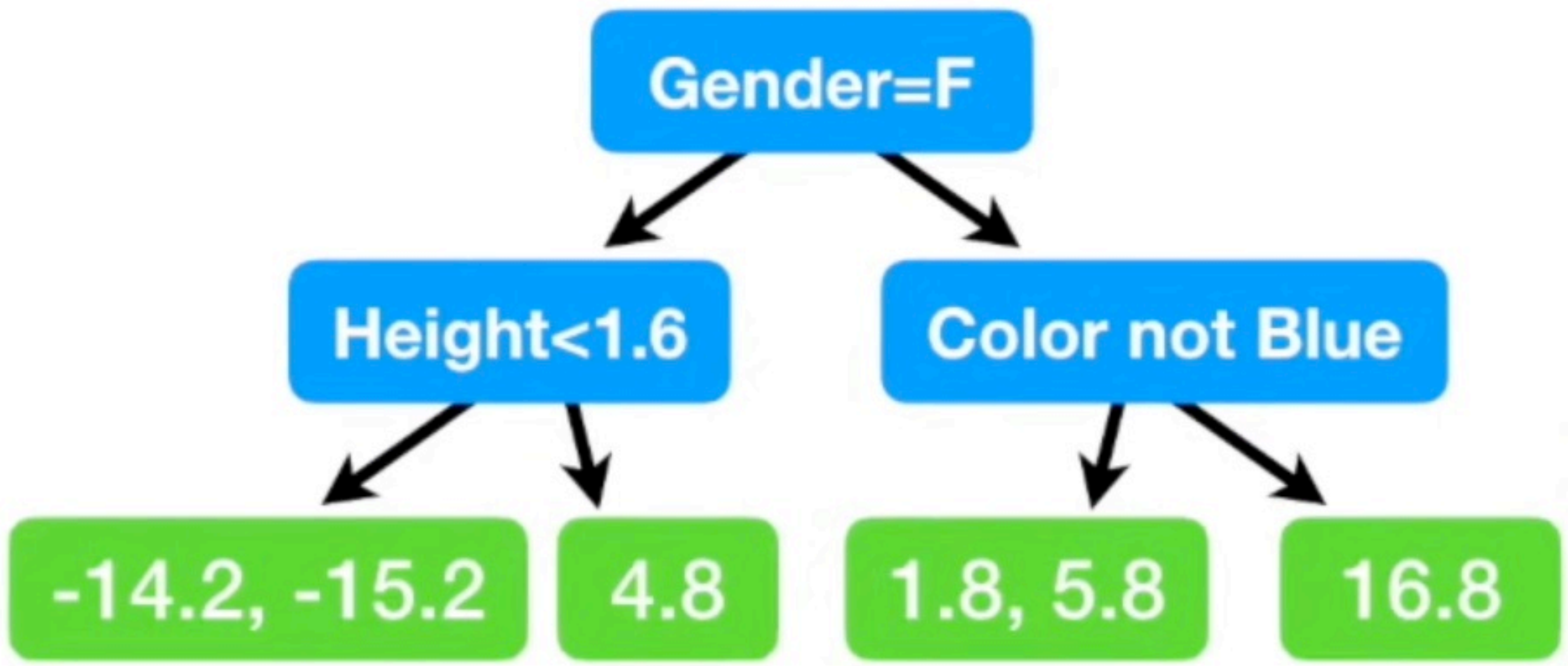
Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
1.5	Blue	Female	56	-15.2
1.8	Red	Male	73	1.8
1.5	Green	Male	77	5.8
1.4	Blue	Female	57	-14.2



Height (m)	Favorite Color	Gender	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

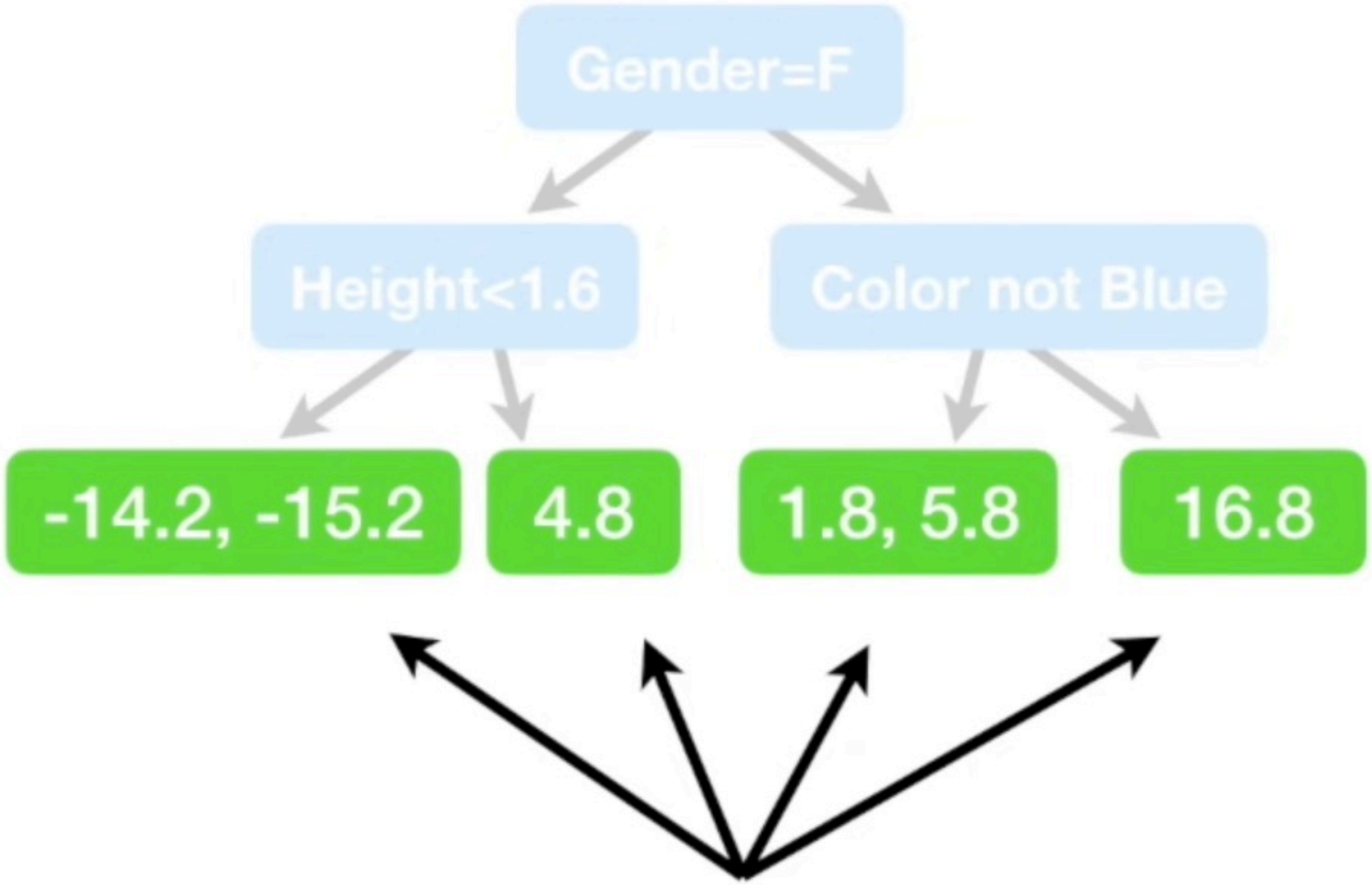
If it seems strange to **Predict** the **Residuals** instead of the original **Weights**, just bear with me and soon all will become clear.

Height (m)	Favorite Color	Gender	[illegible]	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



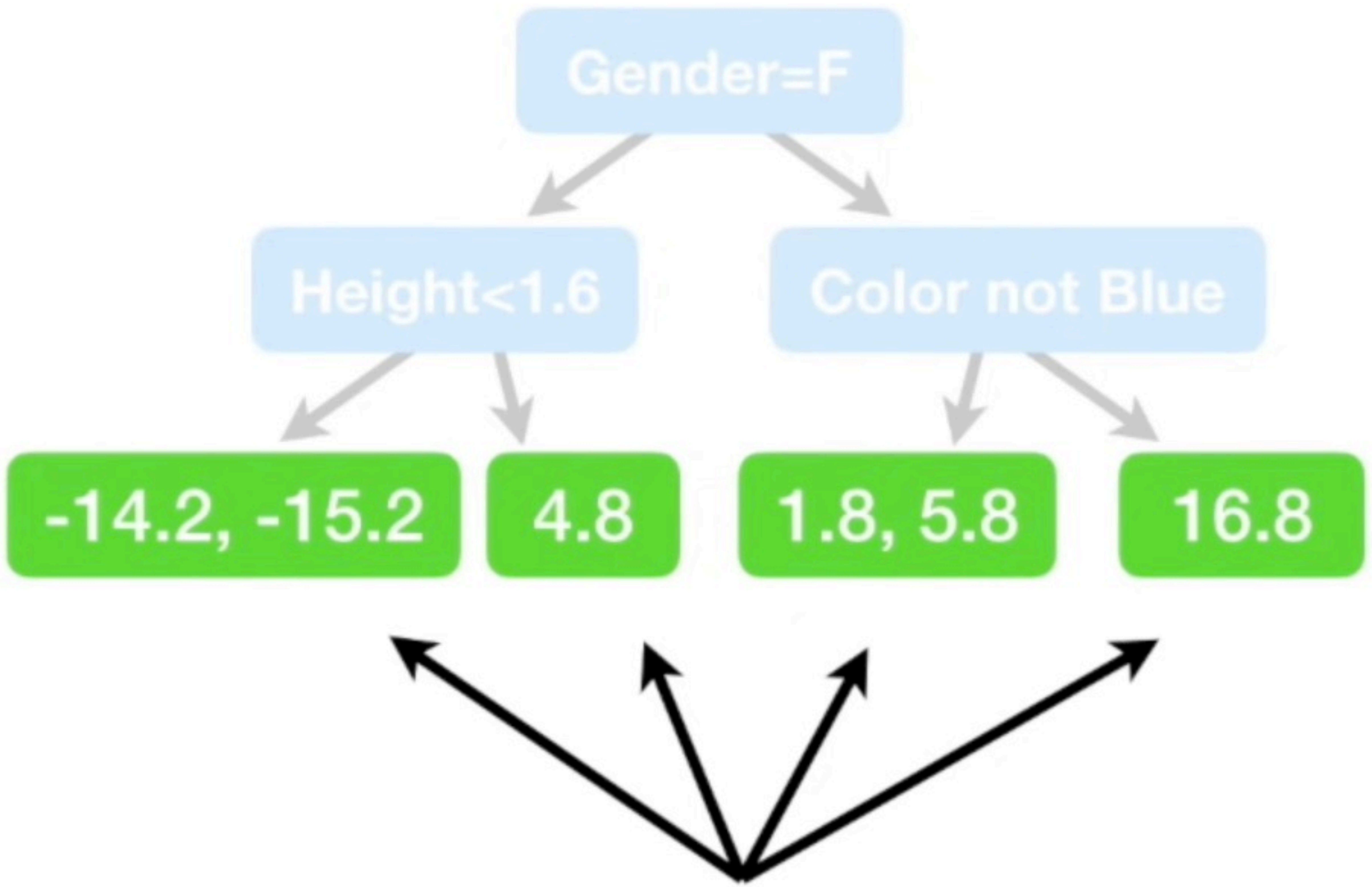
So, setting aside the reason why we are building a tree to **Predict** the **Residuals** for the time being, here's the tree!

Height (m)	Favorite Color	Gender	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



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

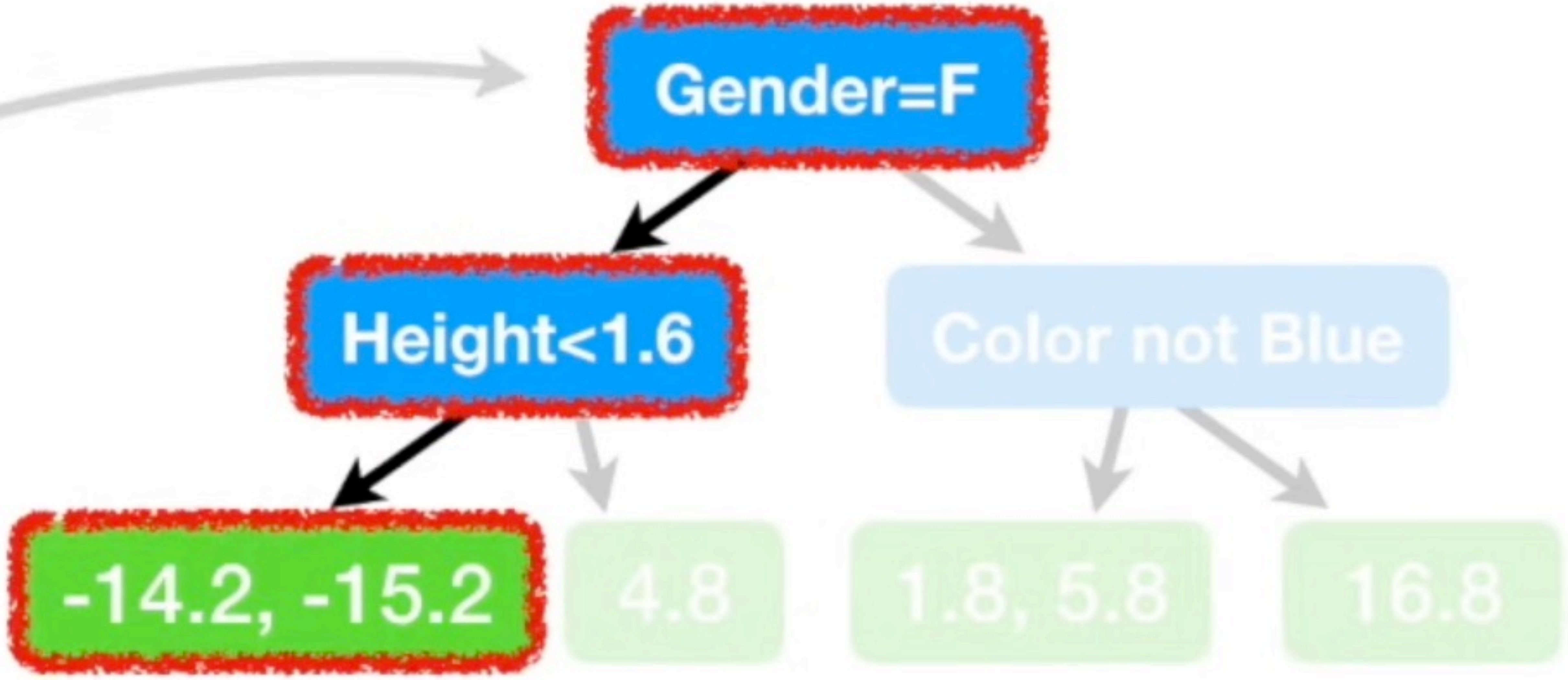
Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
1.5	Blue	Female	56	-15.2
1.8	Red	Male	73	1.8
1.5	Green	Male	77	5.8
1.4	Blue	Female	57	-14.2



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	Weight (kg)	Residual
1.6	Blue	Male	[illegible]	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

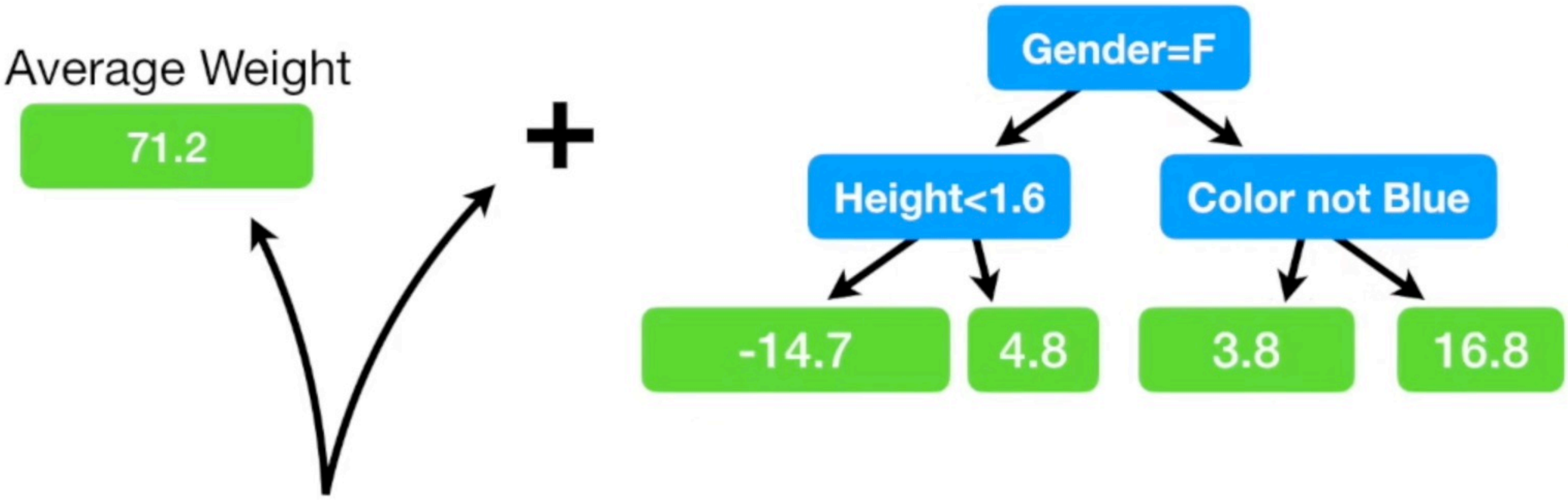


As a result, these two rows of data go to the same leaf.

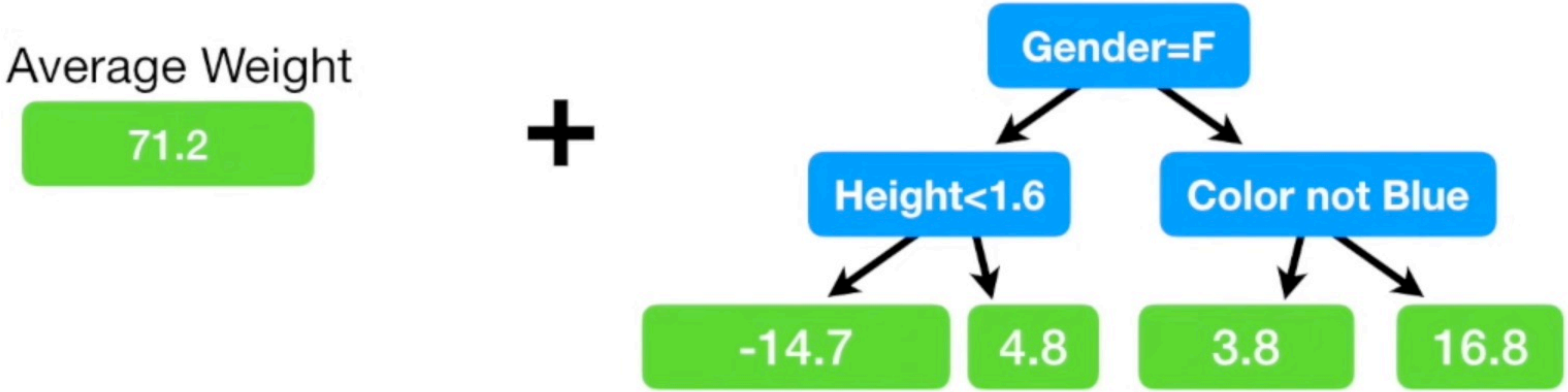
Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	56	16.8
1.6	Green	Female	76	4.8
1.5	Blue	Female	56	-15.2
1.8	Red	Male	73	1.8
1.5	Green	Male	77	5.8
1.4	Blue	Female	57	-14.2

So we replace these residuals with their average.

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



Now we can now combine
the original leaf…



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**.

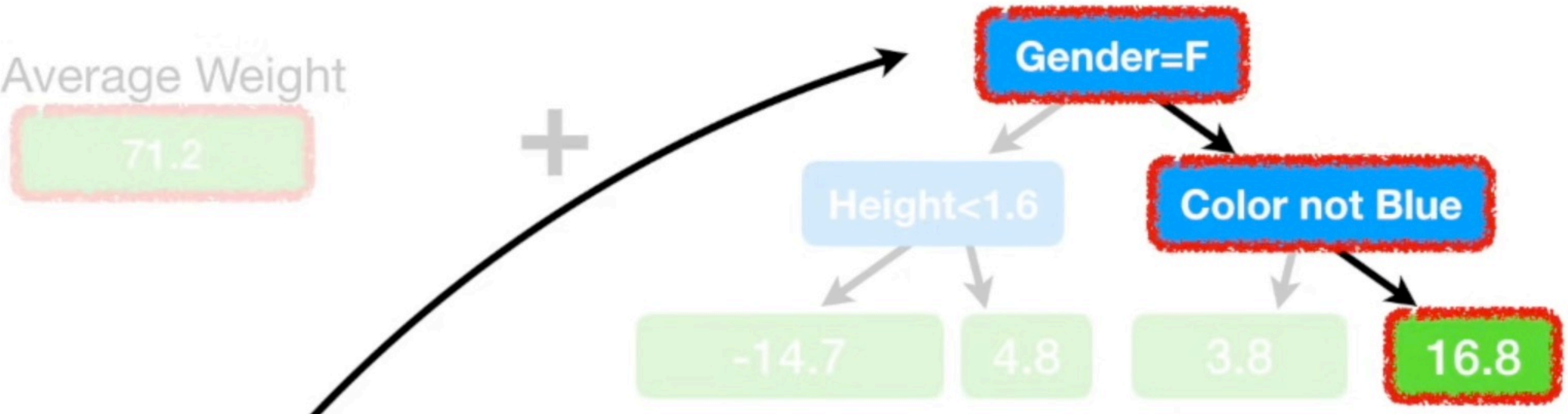
Average Weight

71.2

+

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

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



…then we run the data down the tree…

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

…and we get **16.8**…

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

…so the **Predicted Weight** = 71.2 + 16.8 = 88

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

Predicted Weight = 71.2 + 16.8 $\boxed{= 88}$

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

Is this awesome???

Predicted Weight = 71.2 + 16.8 = 88

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

No. The model fits the **Training Data** too well.

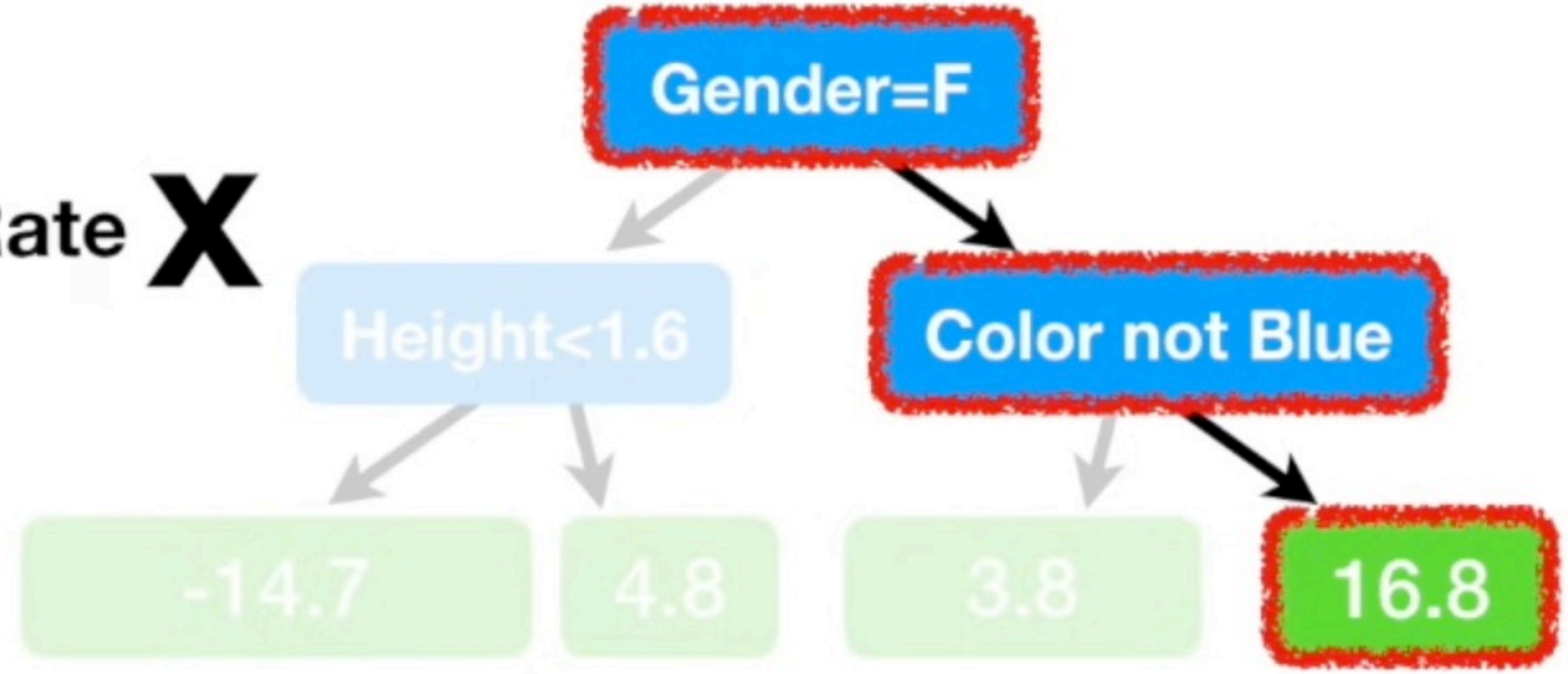
Predicted Weight = 71.2 + 16.8 = 88

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



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

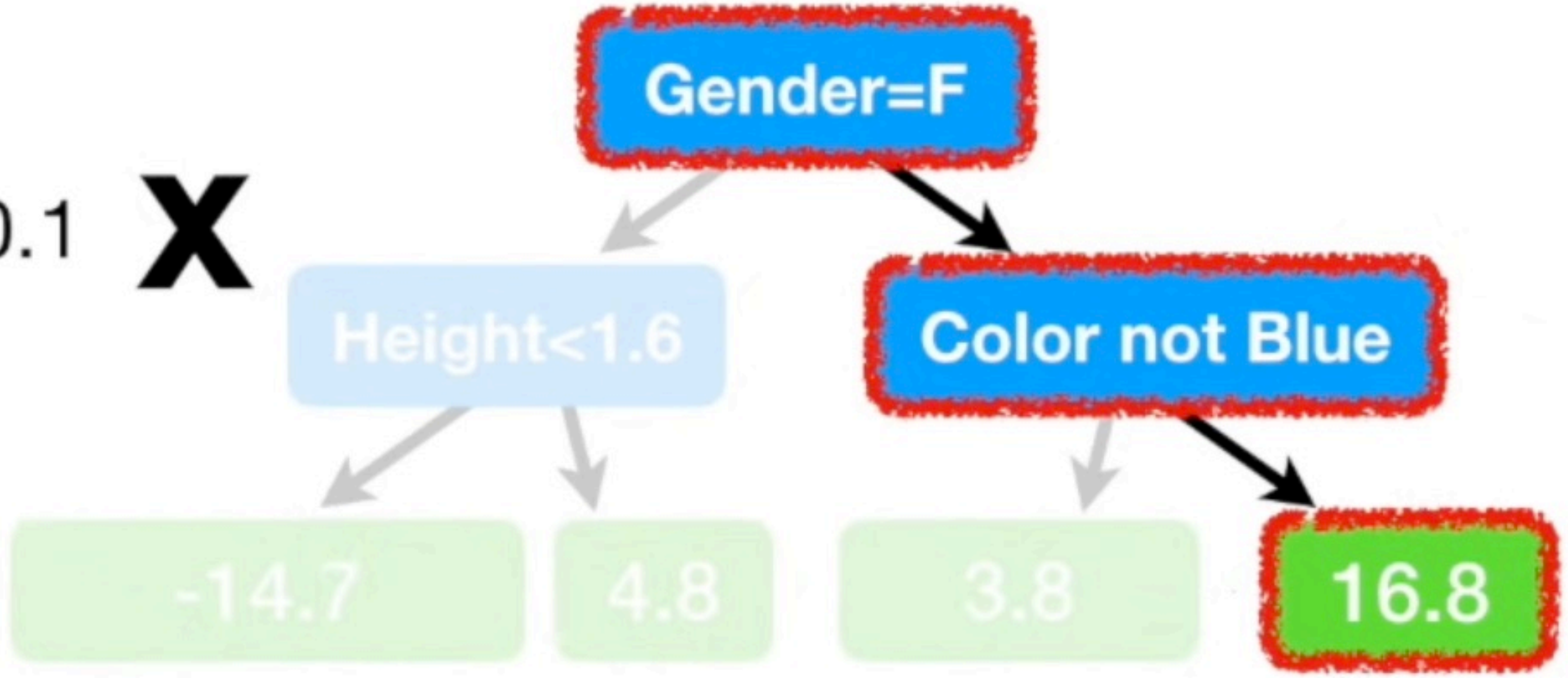
Gradient Boost deals with this problem by using a **Learning Rate** to scale the contribution from the new tree.

The **Learning Rate** is a value between **0** and **1**.

Average Weight

71.2

+ 0.1 X



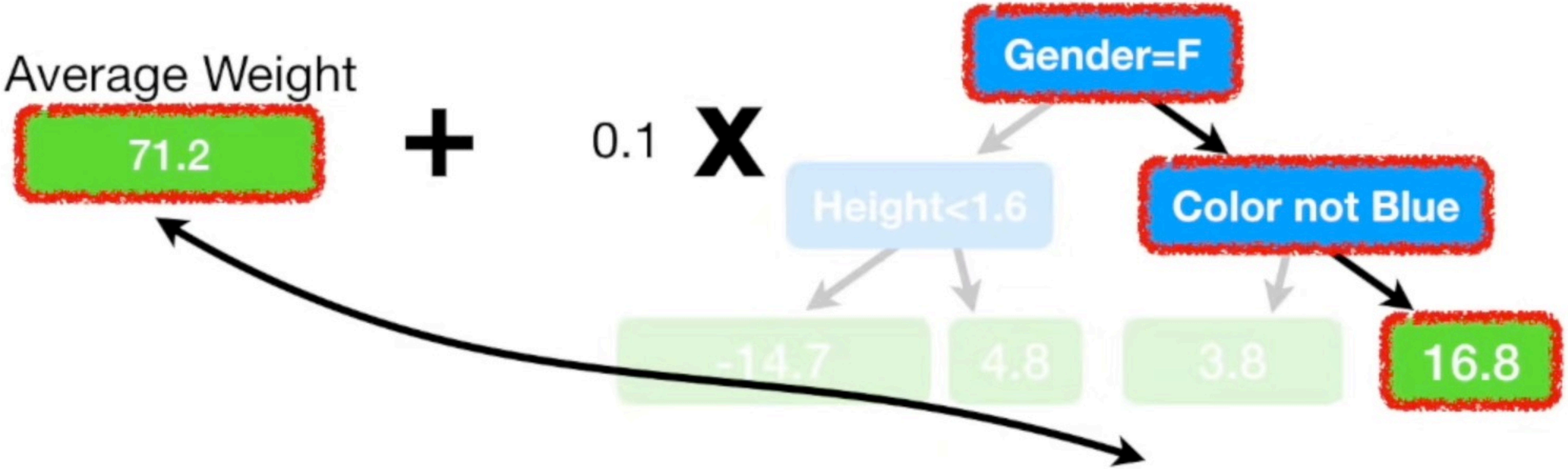
Now the **Predicted Weight** = 71.2 + (0.1 × 16.8) = 72.9

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

Predicted Weight = $71.2 + (0.1 \times 16.8) \boxed{= 72.9}$

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

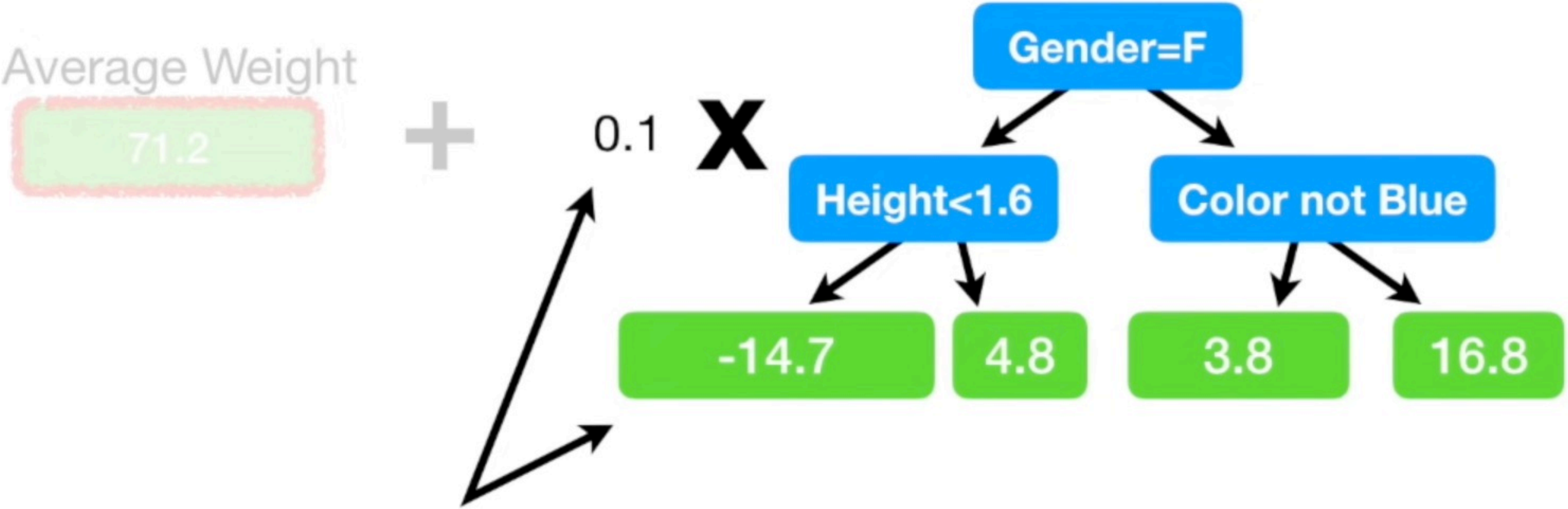
With the **Learning Rate** set to **0.1**, the new **Prediction** isn't as good as as it was before…



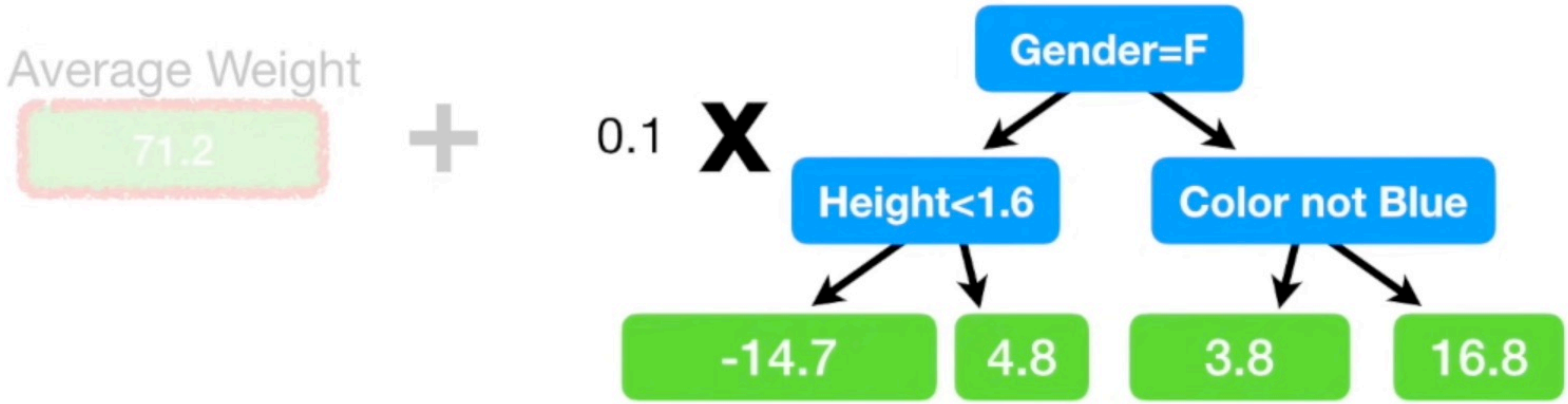
Predicted Weight = 71.2 + (0.1 × 16.8) $\boxed{= 72.9}$

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

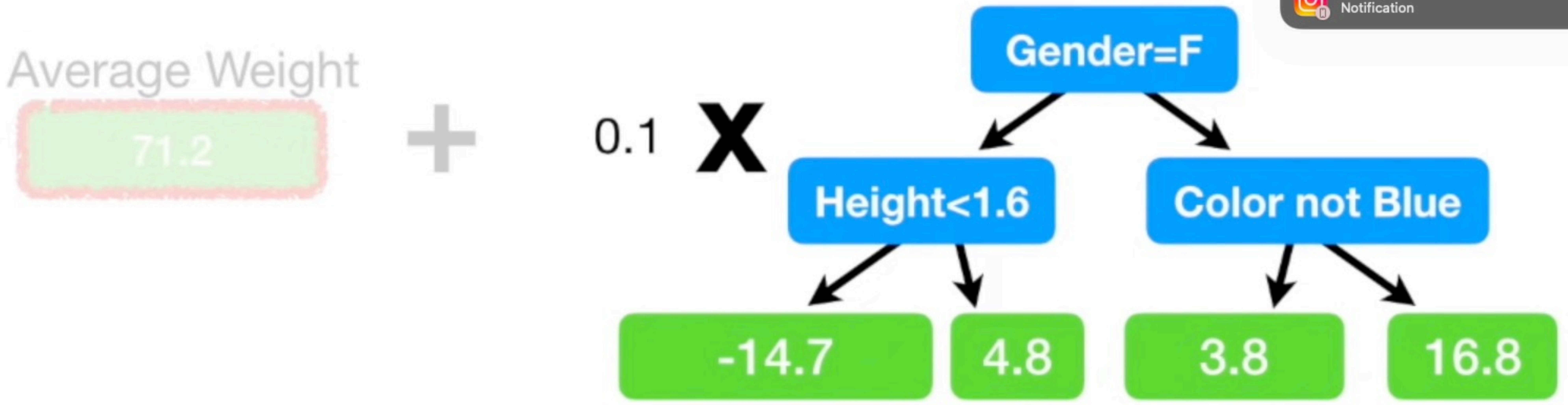
…but it's a little bit better than the **Prediction** made with just the original leaf, which predicted that all samples would weigh **71.2**.



In other words, scaling the tree by the **Learning Rate** results in a small step in the right direction.



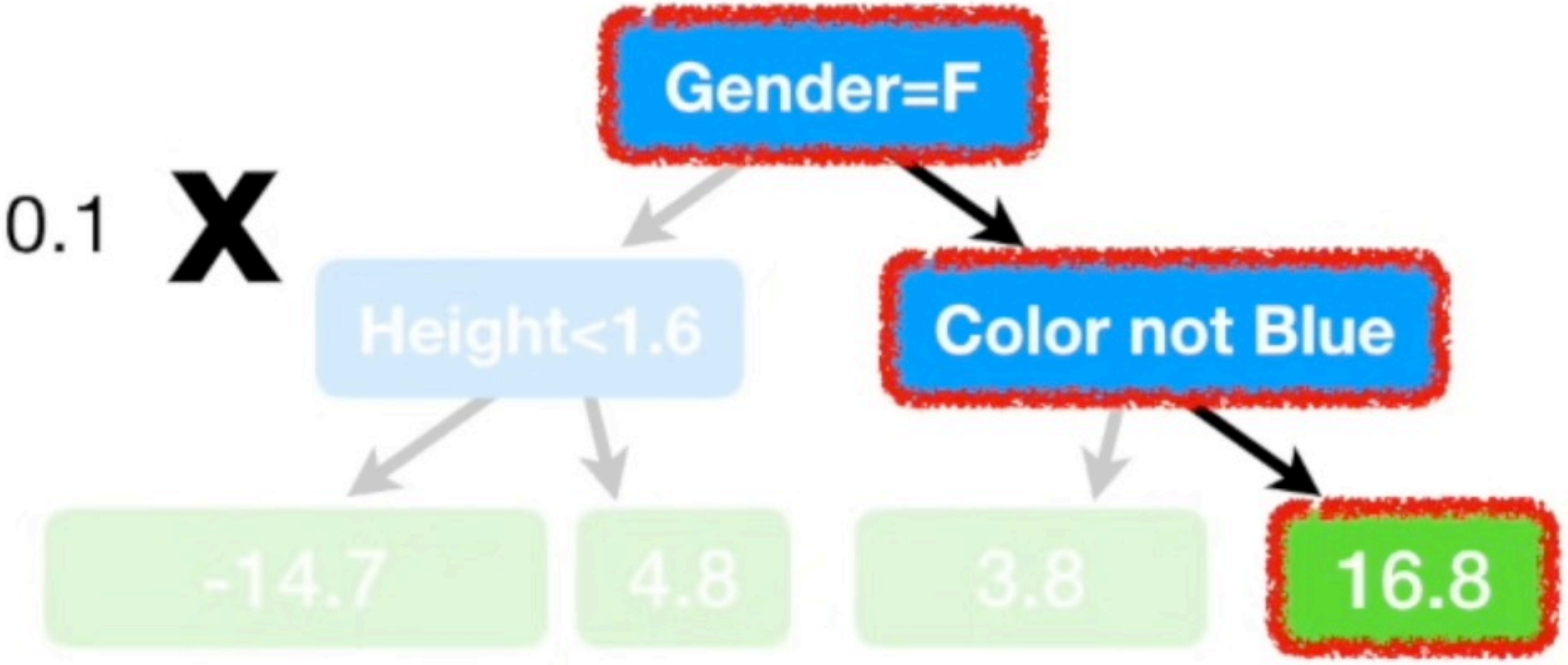
According to the dude that invented **Gradient Boost**, Jerome Friedman, empirical evidence shows that taking lots of small steps in the right direction results in better **Predictions** with a **Testing Dataset**, i.e. lower **Variance**.



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

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	

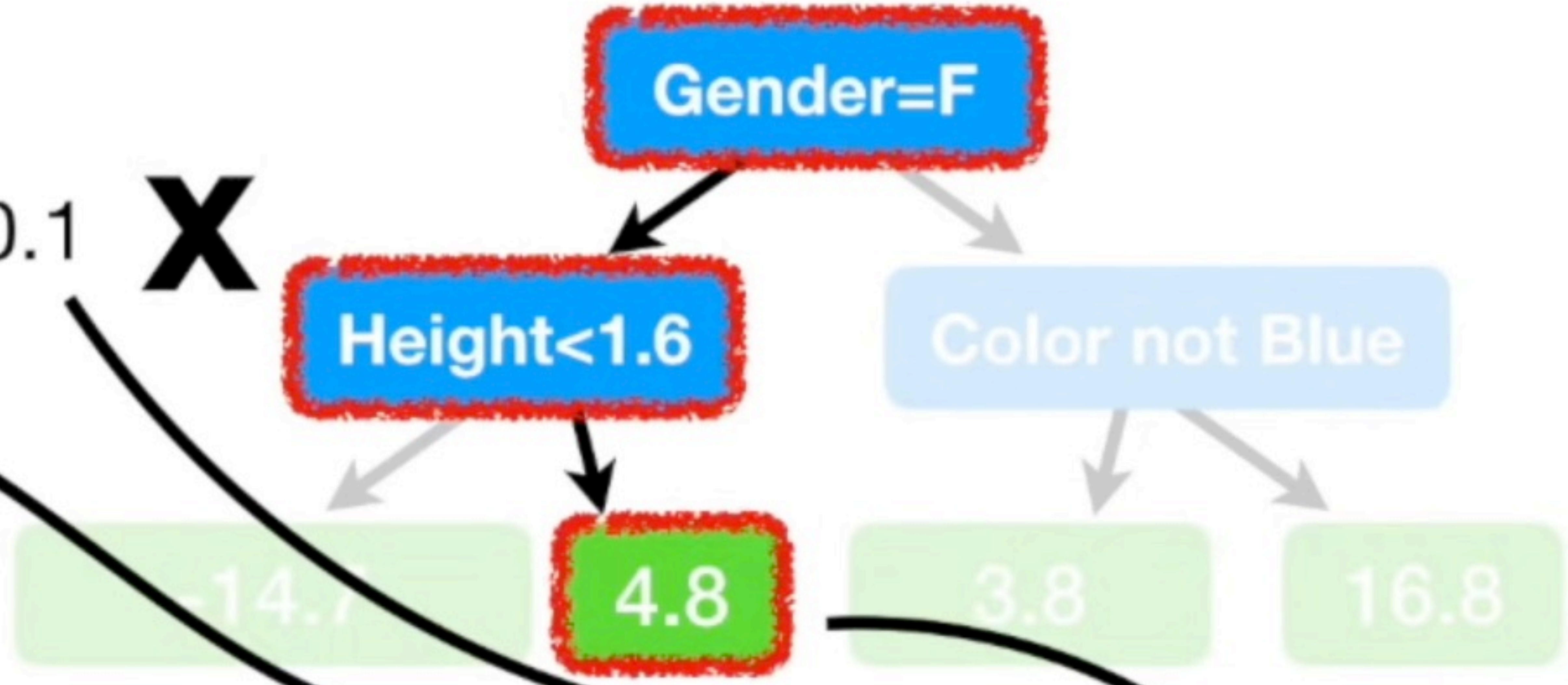
Residual = (**88** - (71.2 + 0.1 × 16.8))

$\boxed{= 15.1}$

…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	
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))

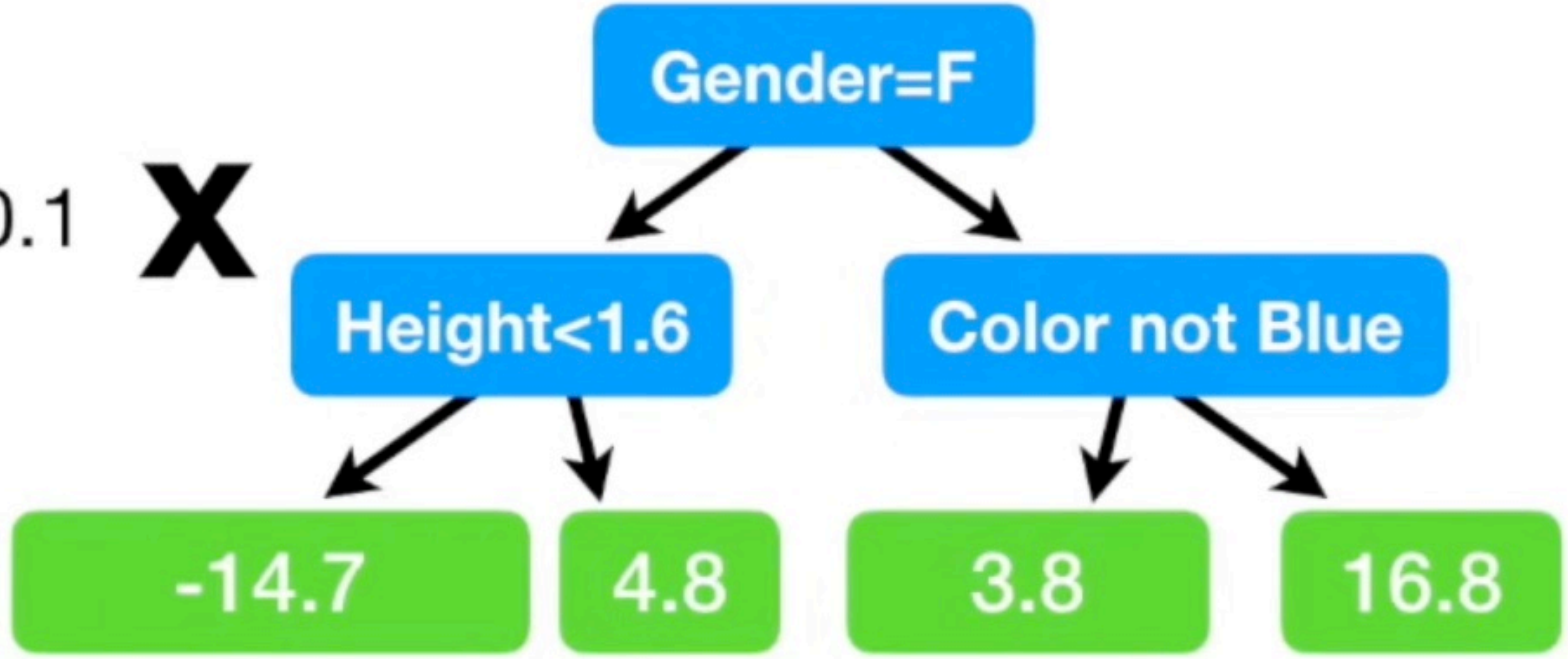
Then we repeat for the all of the other individuals in the **Training Dataset**.

Average Weight

71.2

+

0.1 X



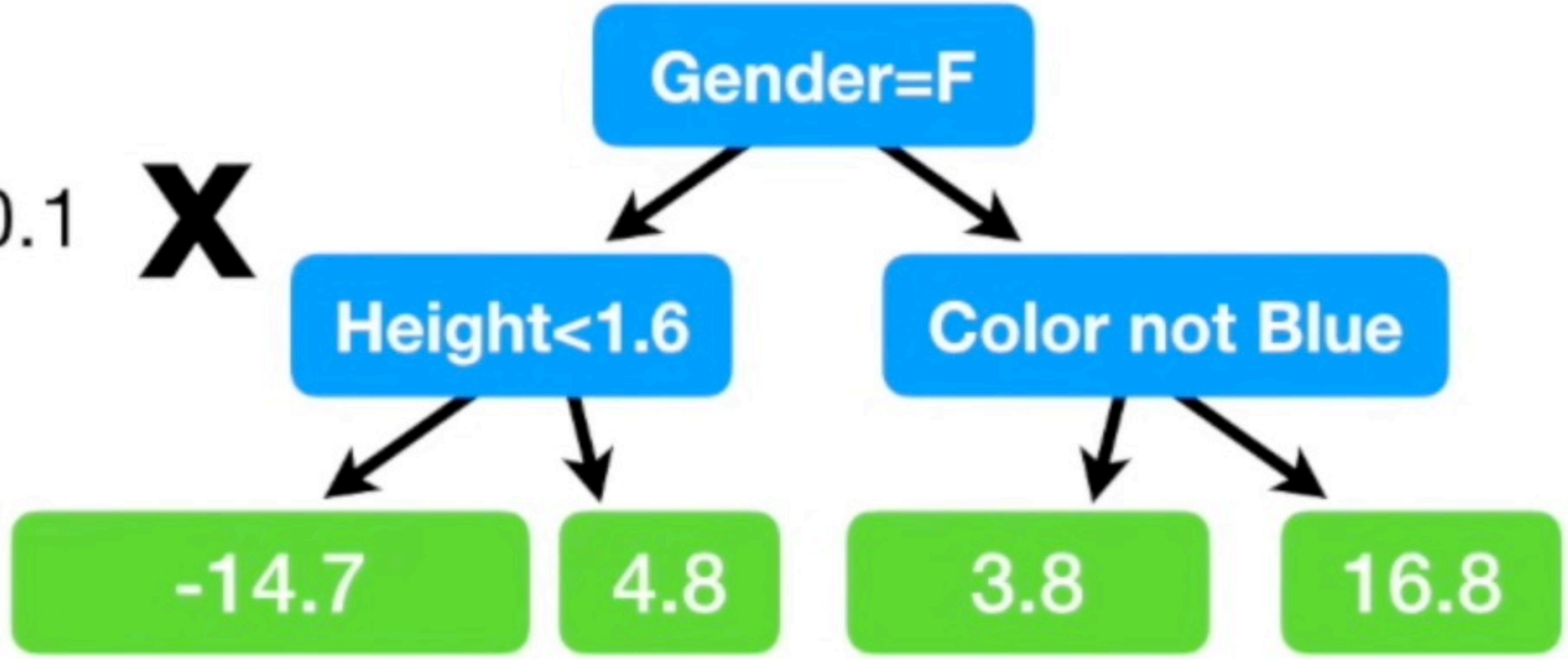
Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	15.1
1.6	Green	Female	76	4.3
1.5	Blue	Female	56	-13.7
1.8	Red	Male	73	1.4
1.5	Green	Male	77	5.4
1.4	Blue	Female	57	-12.7

Small bam.

Average Weight

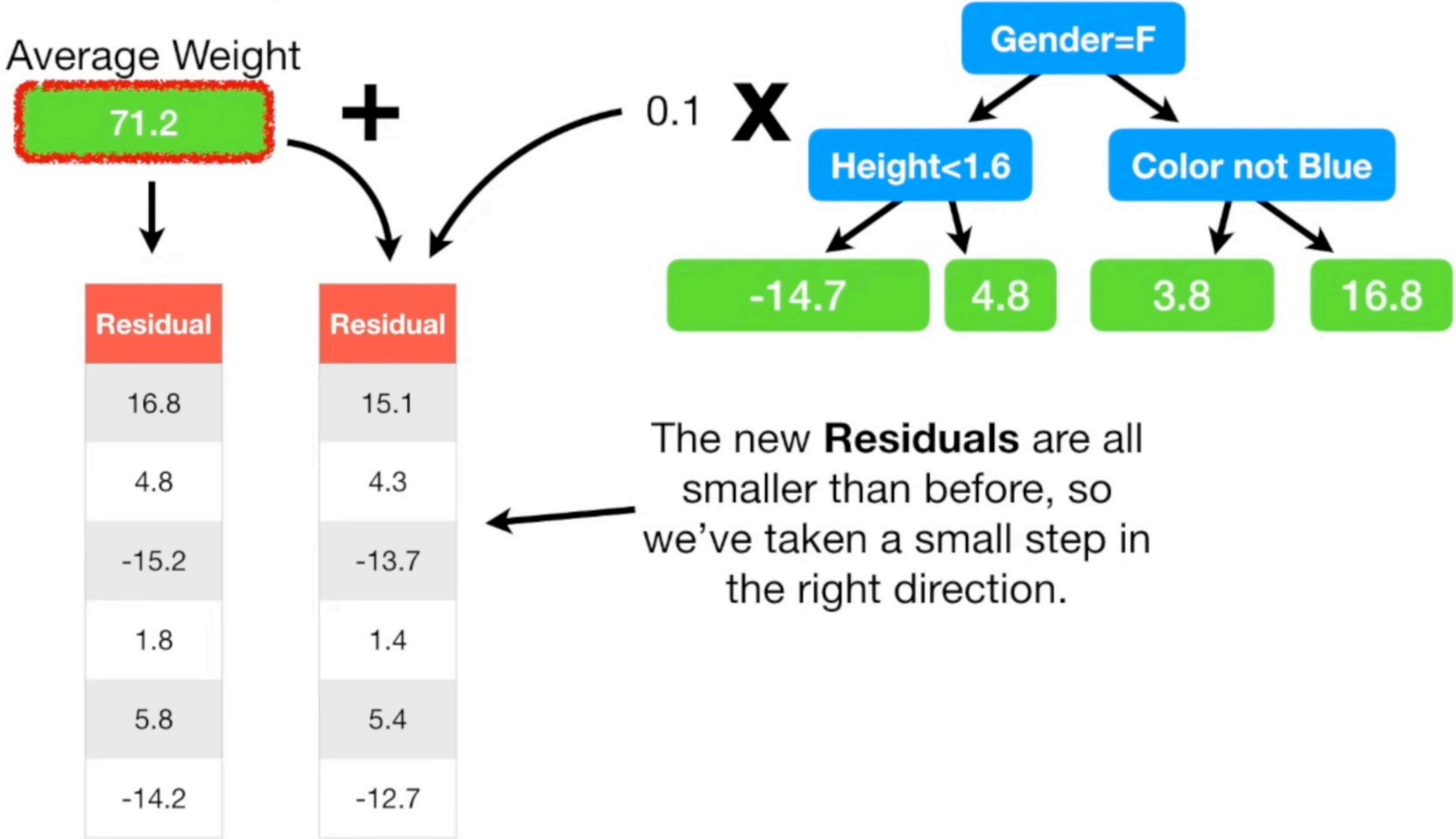
71.2

+ 0.1 x

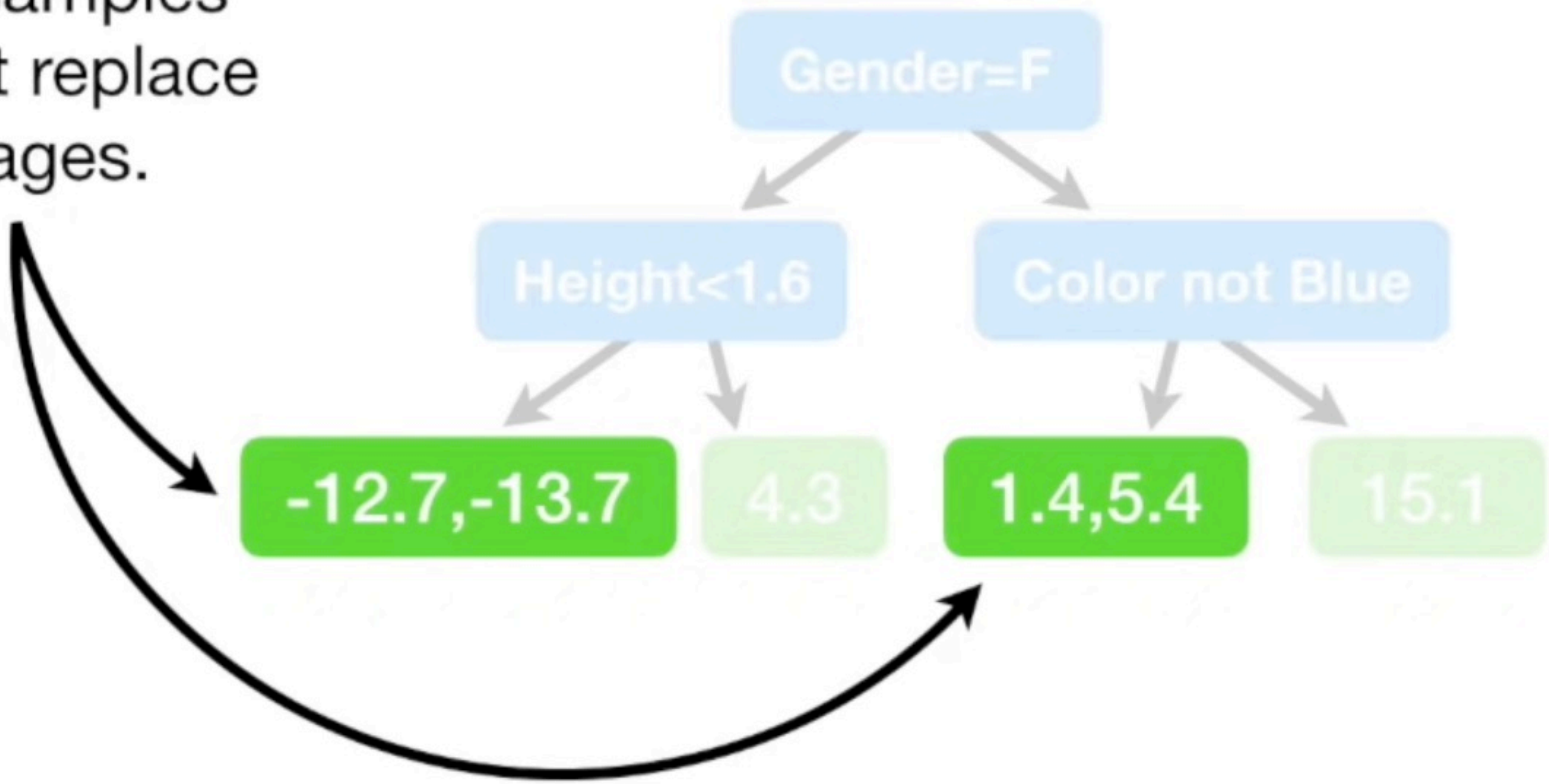


Residual	Residual
16.8	15.1
4.8	4.3
-15.2	-13.7
1.8	1.4
5.8	5.4
-14.2	-12.7

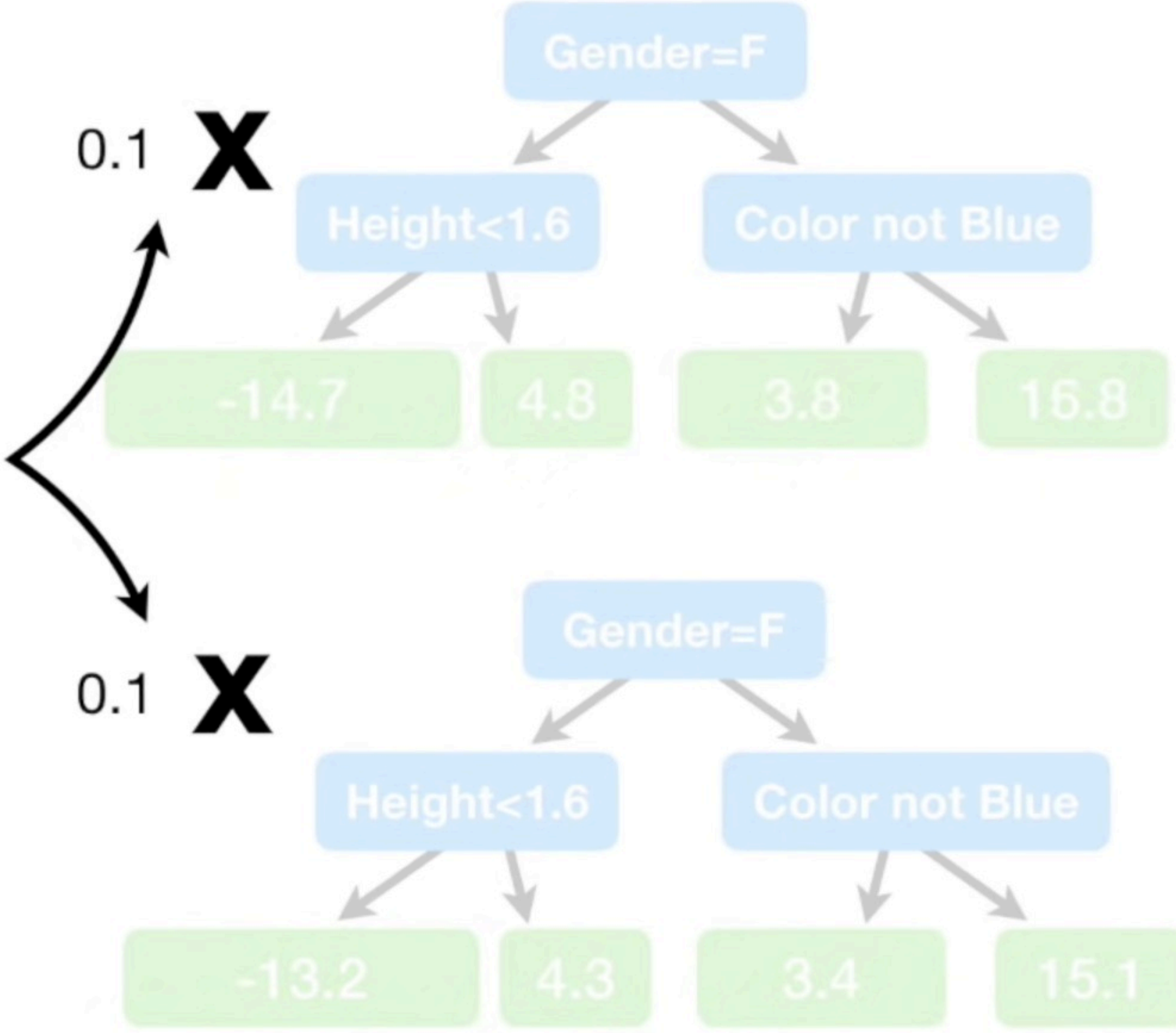
...and these are the **Residuals** after adding the new tree scaled by the **Learning Rate**.

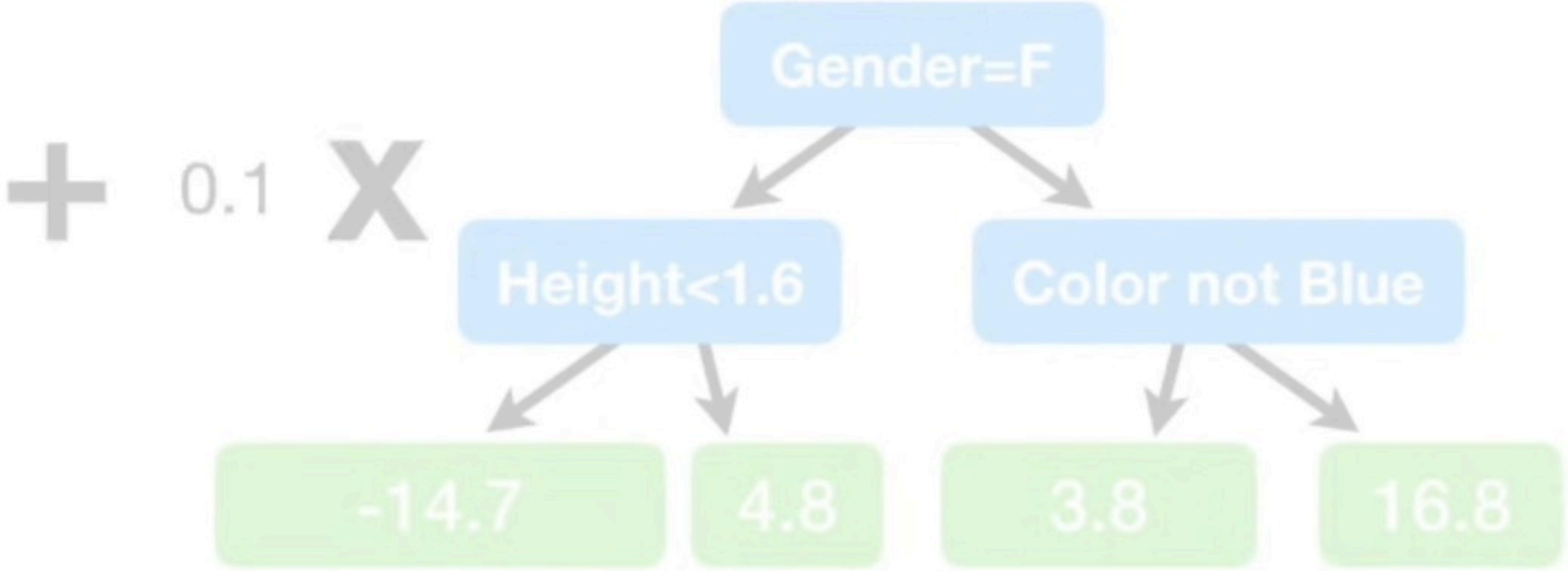

Average Weight
71.2
+
0.1
X
Gender=F
Height<1.6
Color not Blue
-14.7
4.8
3.8
16.8
Residual
16.8
4.8
-15.2
1.8
5.8
-14.2
Residual
15.1
4.3
-13.7
1.4
5.4
-12.7
The new **Residuals** are all smaller than before, so we've taken a small step in the right direction.

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

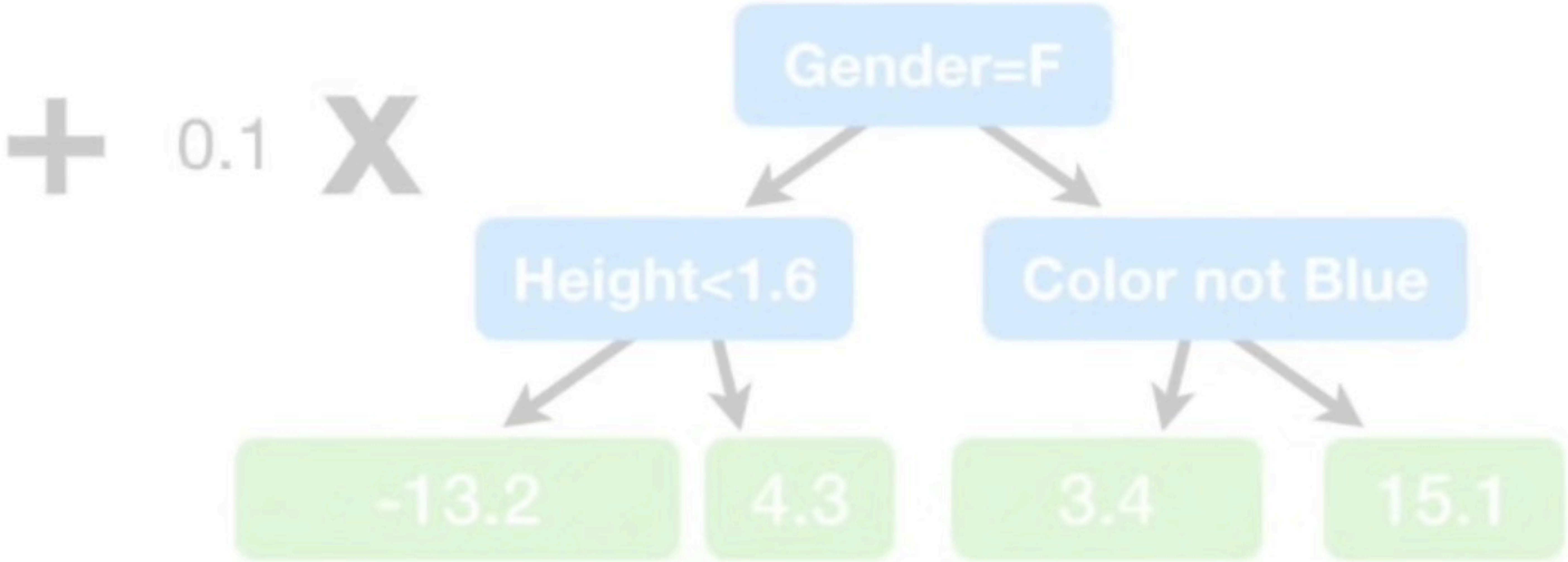


NOTE: We scale all of the **Trees** by the **Learning Rate**, which we set to **0.1**…

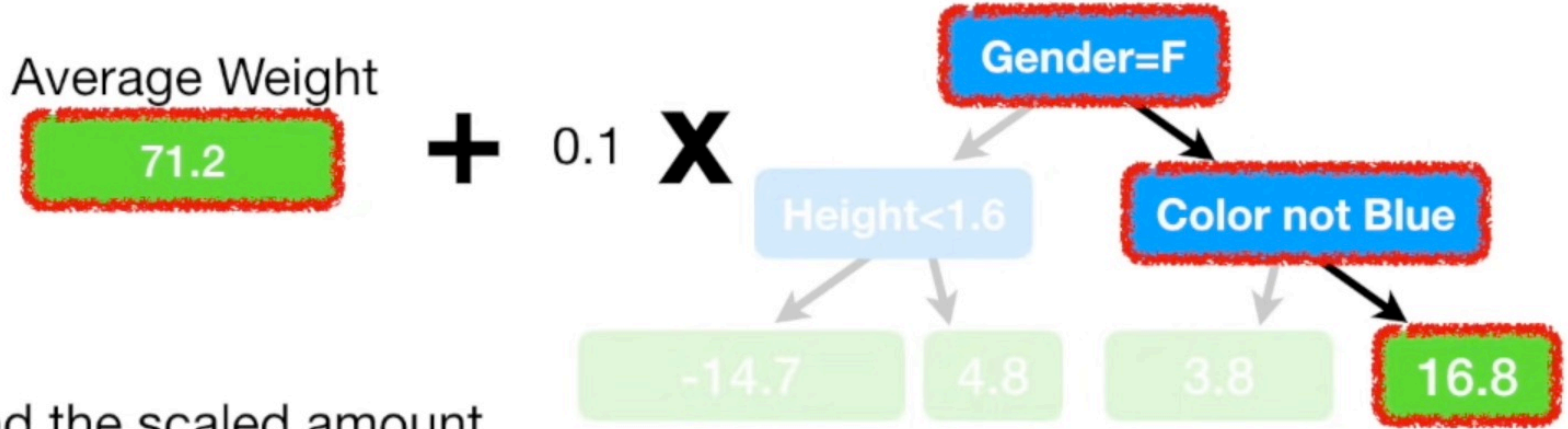




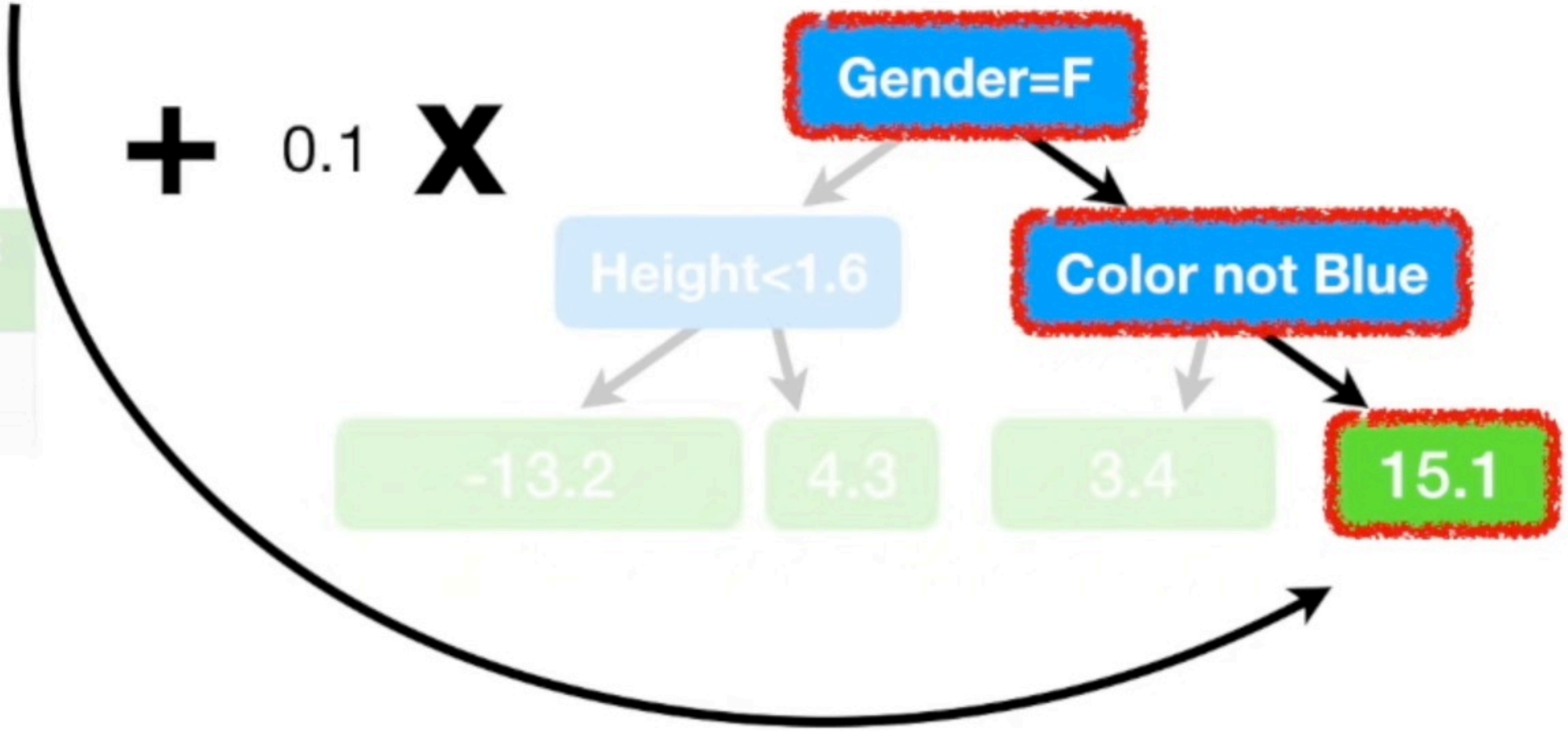
Now we're ready to make a new **Prediction** from the **Training Data**.



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



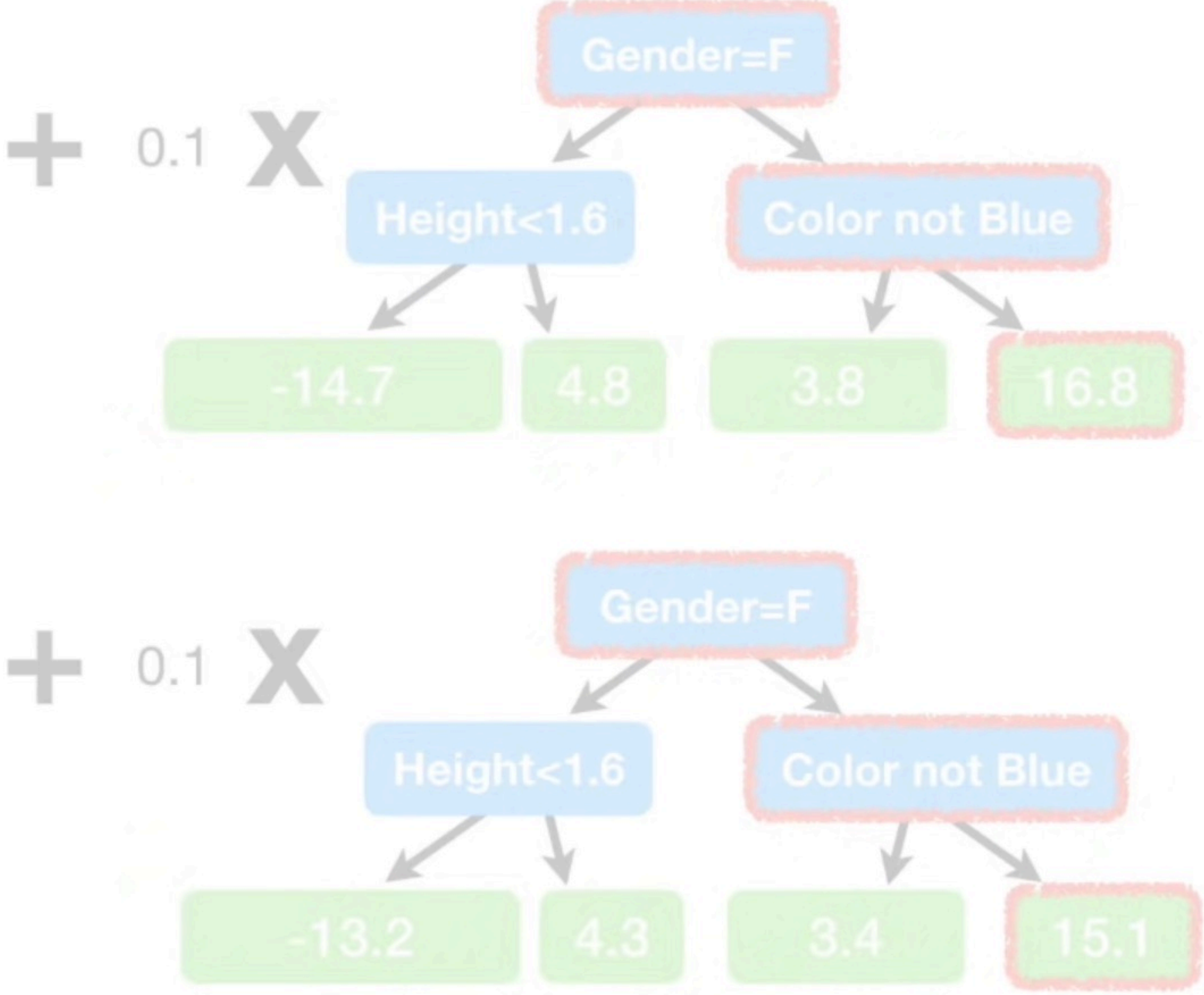
...and the scaled amount
from the second **Tree**.



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

Average Weight

71.2

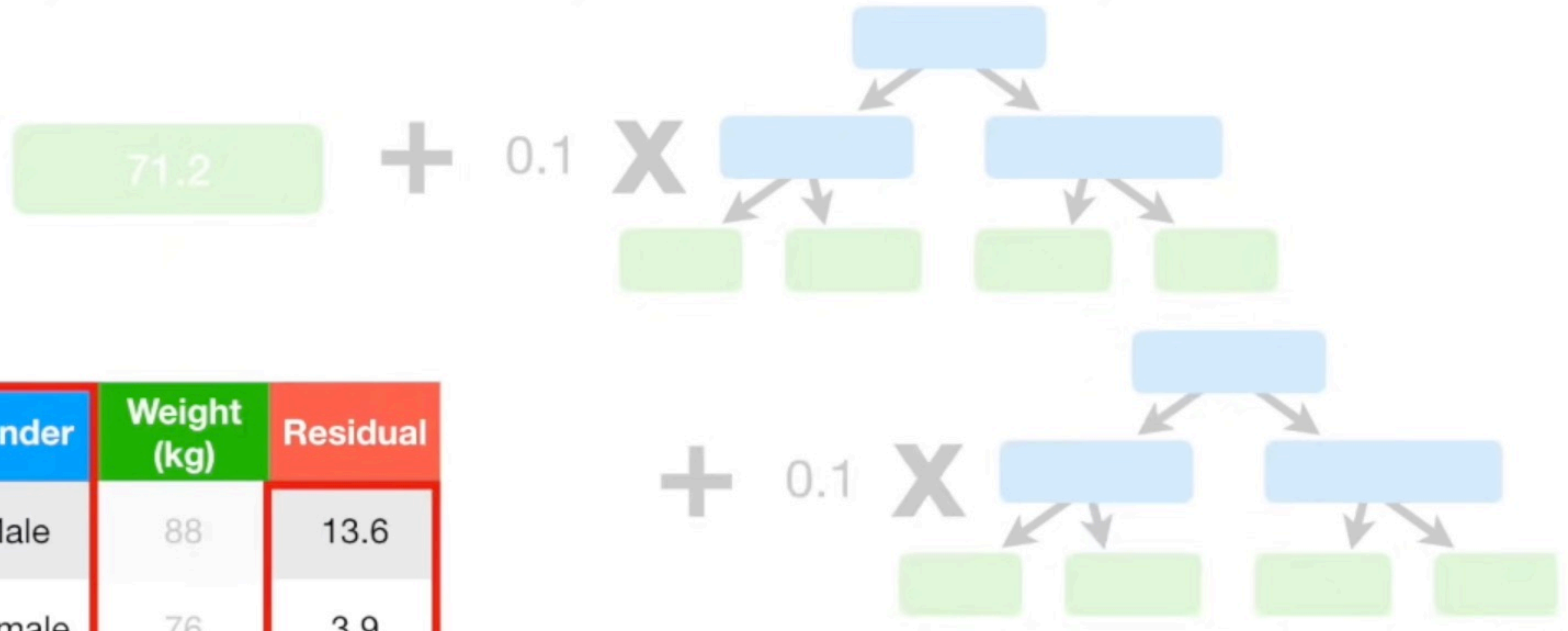


Which is another small step closer to the **Observed Weight**.

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

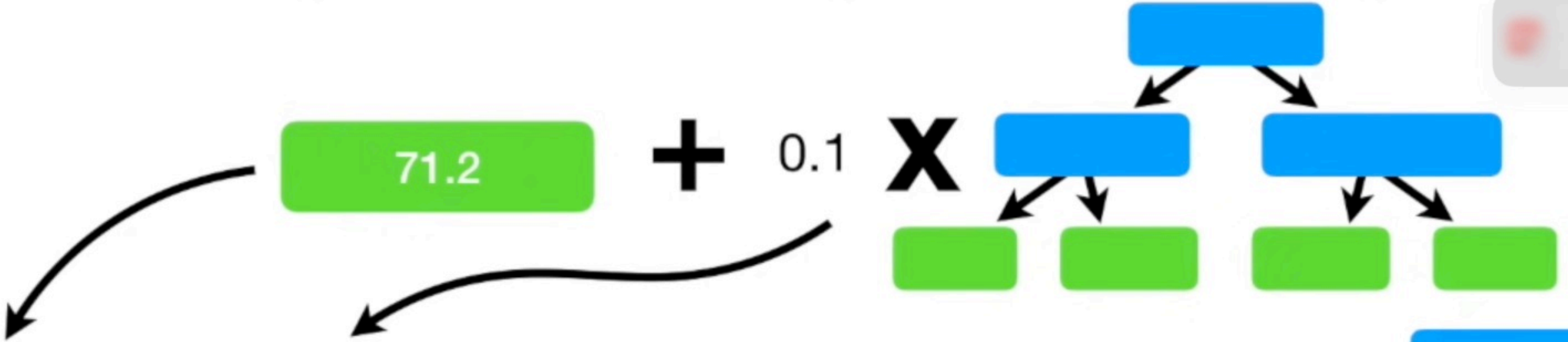
$$\boxed{= 74.4}$$

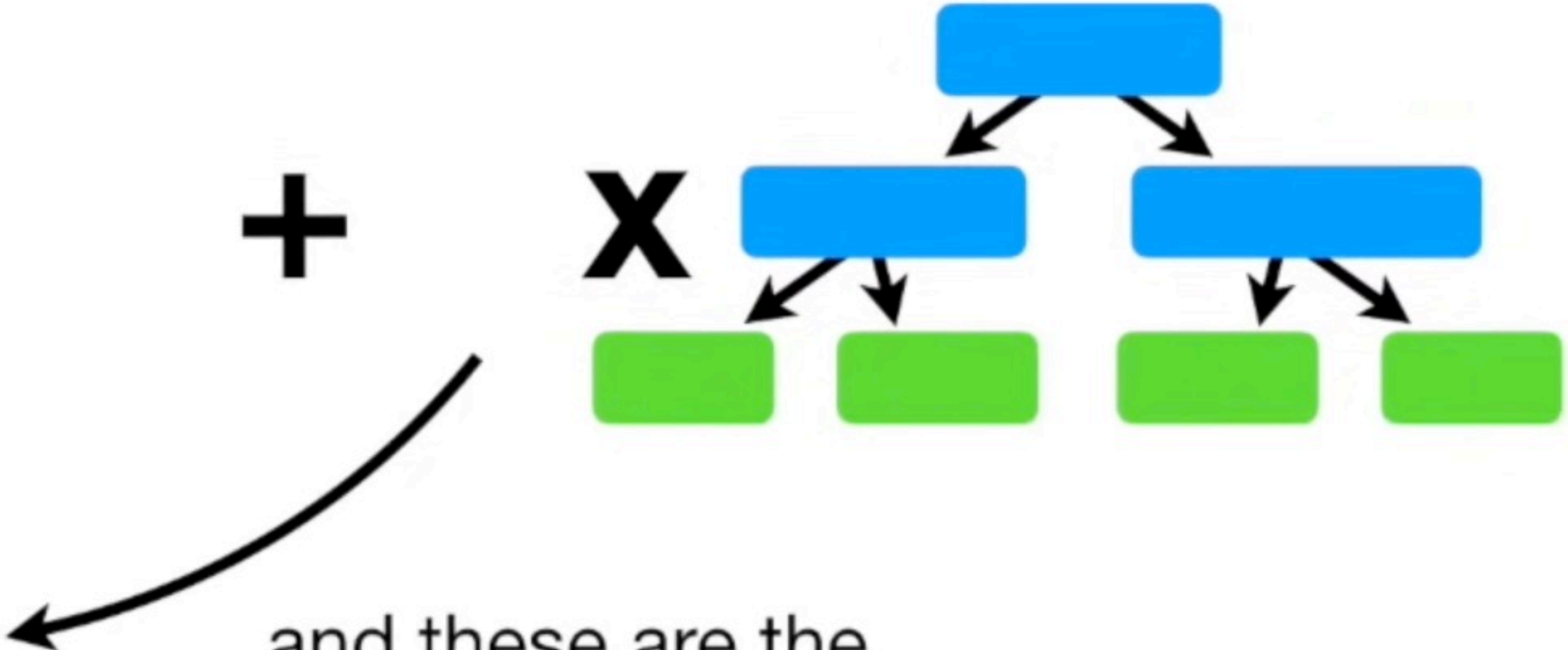
Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88



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

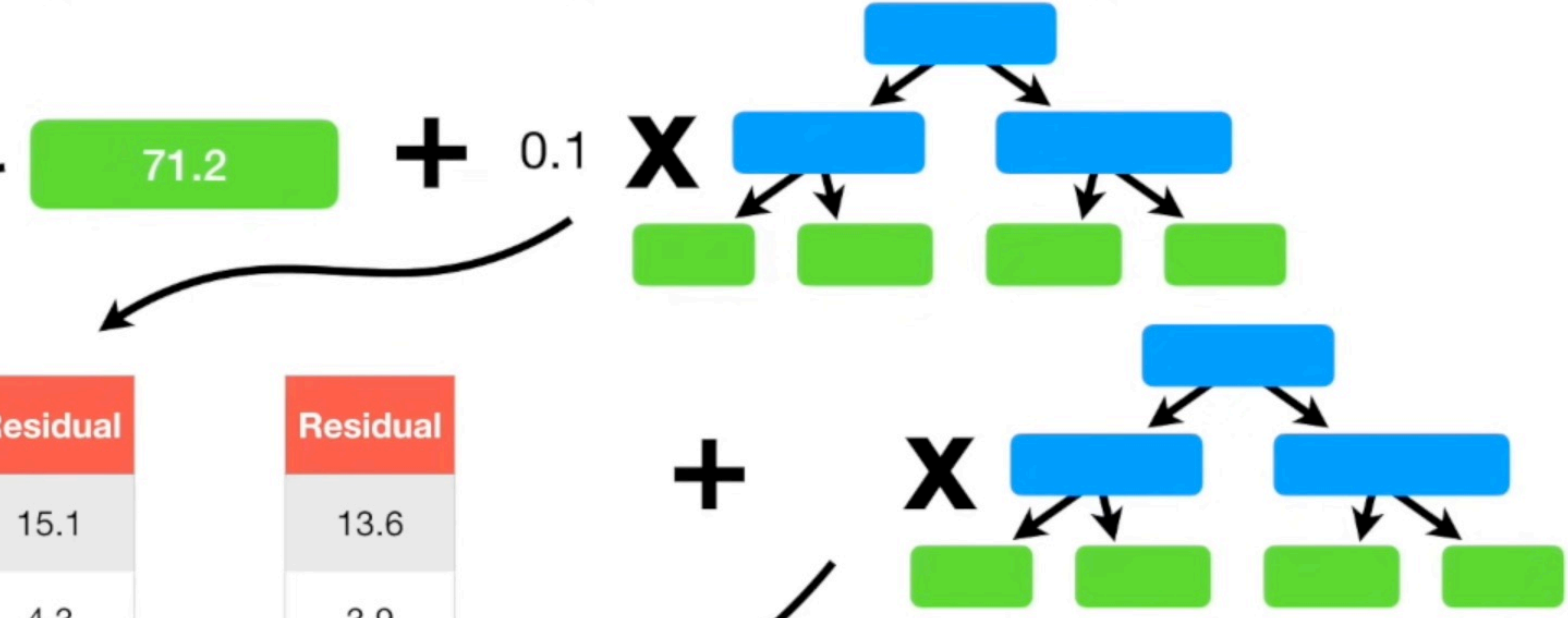
…to calculate new **Residuals**.





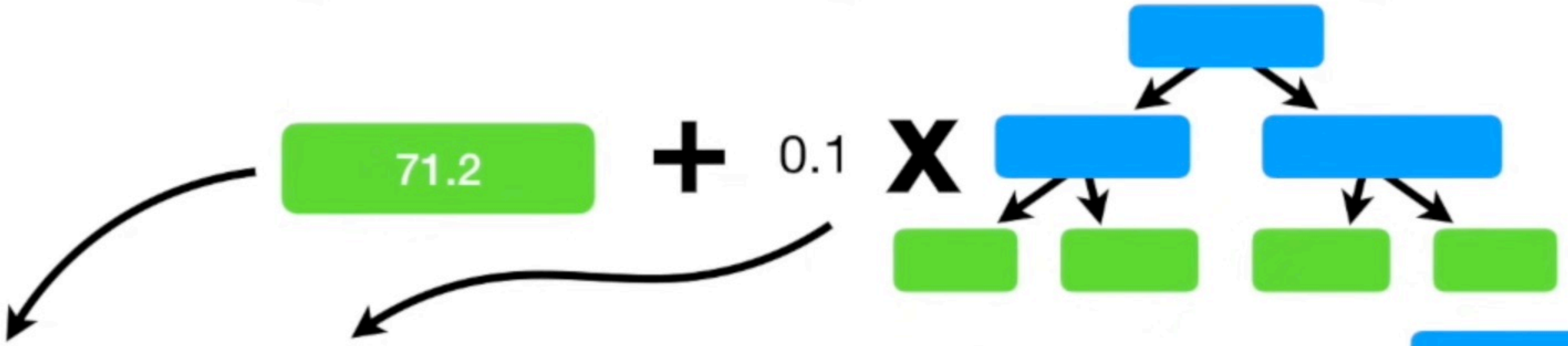
Residual	Residual	Residual
16.8	15.1	13.6
4.8	4.3	3.9
-15.2	-13.7	-12.4
1.8	1.4	1.1
5.8	5.4	5.1
-14.2	-12.7	-11.4

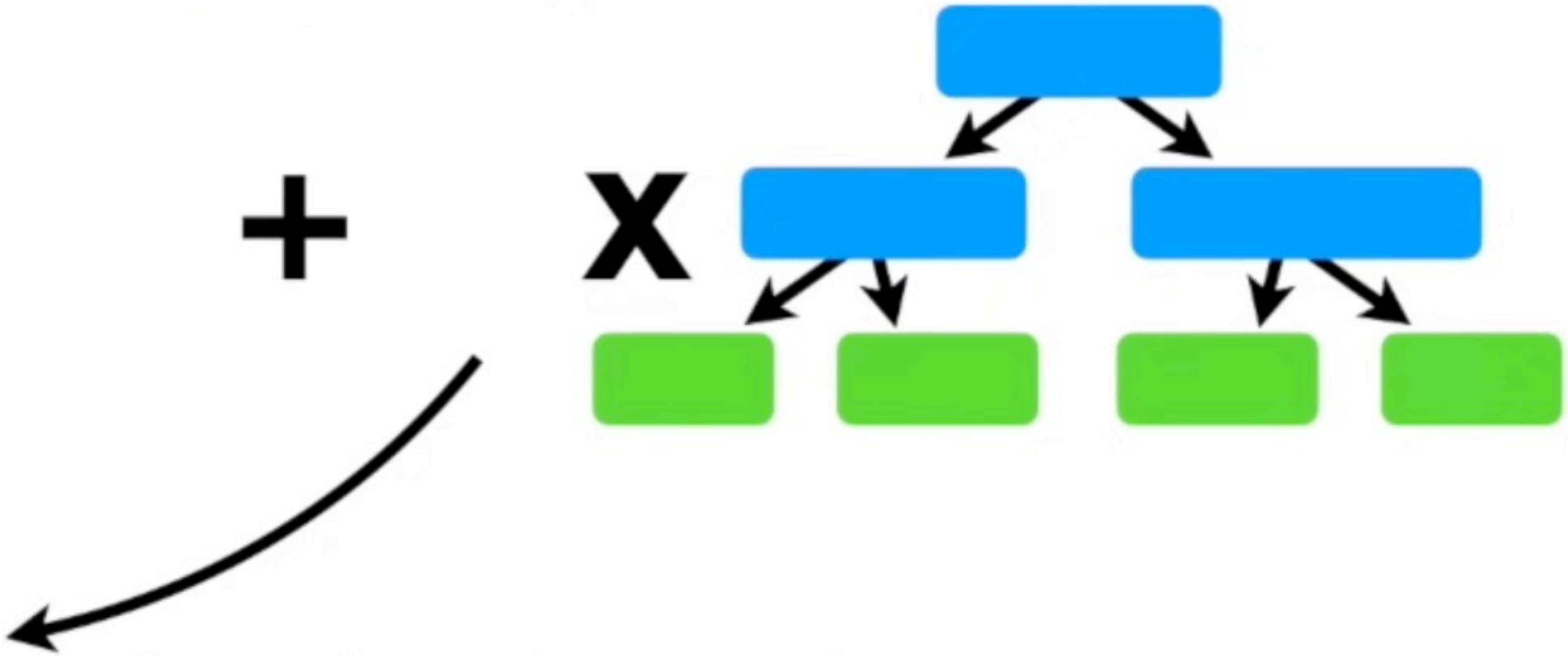
…and these are the **Residuals** after we added the second **Tree** to the **Prediction**…



Residual	Residual	Residual
16.8	15.1	13.6
4.8	4.3	3.9
-15.2	-13.7	-12.4
1.8	1.4	1.1
5.8	5.4	5.1
-14.2	-12.7	-11.4

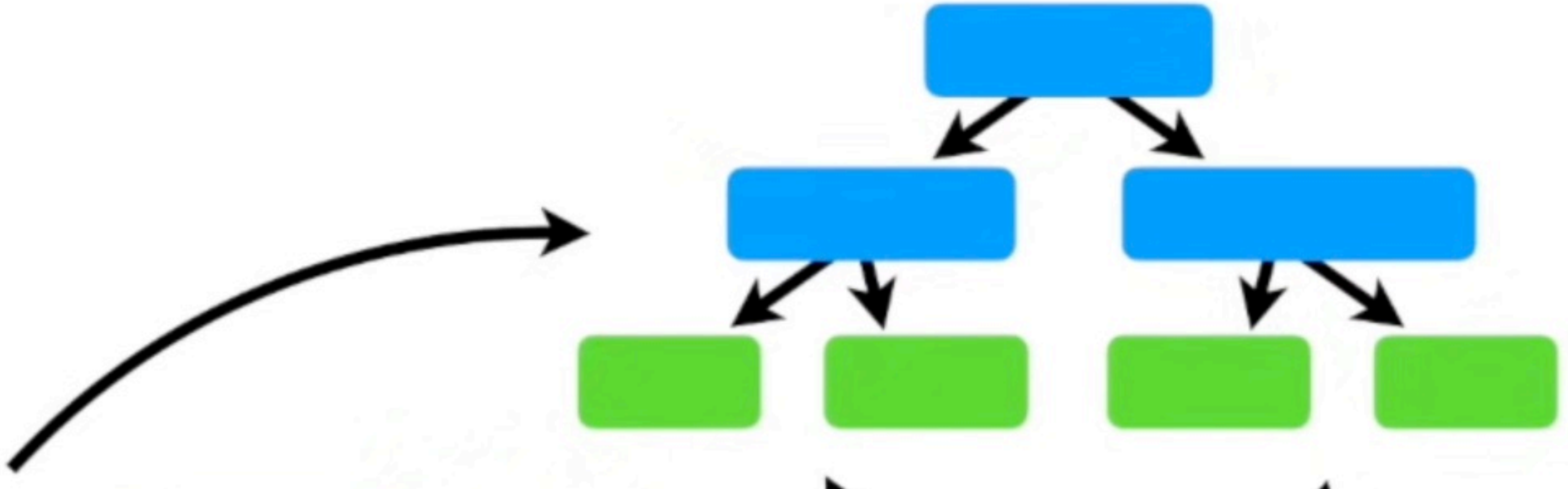
Each time we add a tree to the **Prediction**, the **Residuals** get smaller.





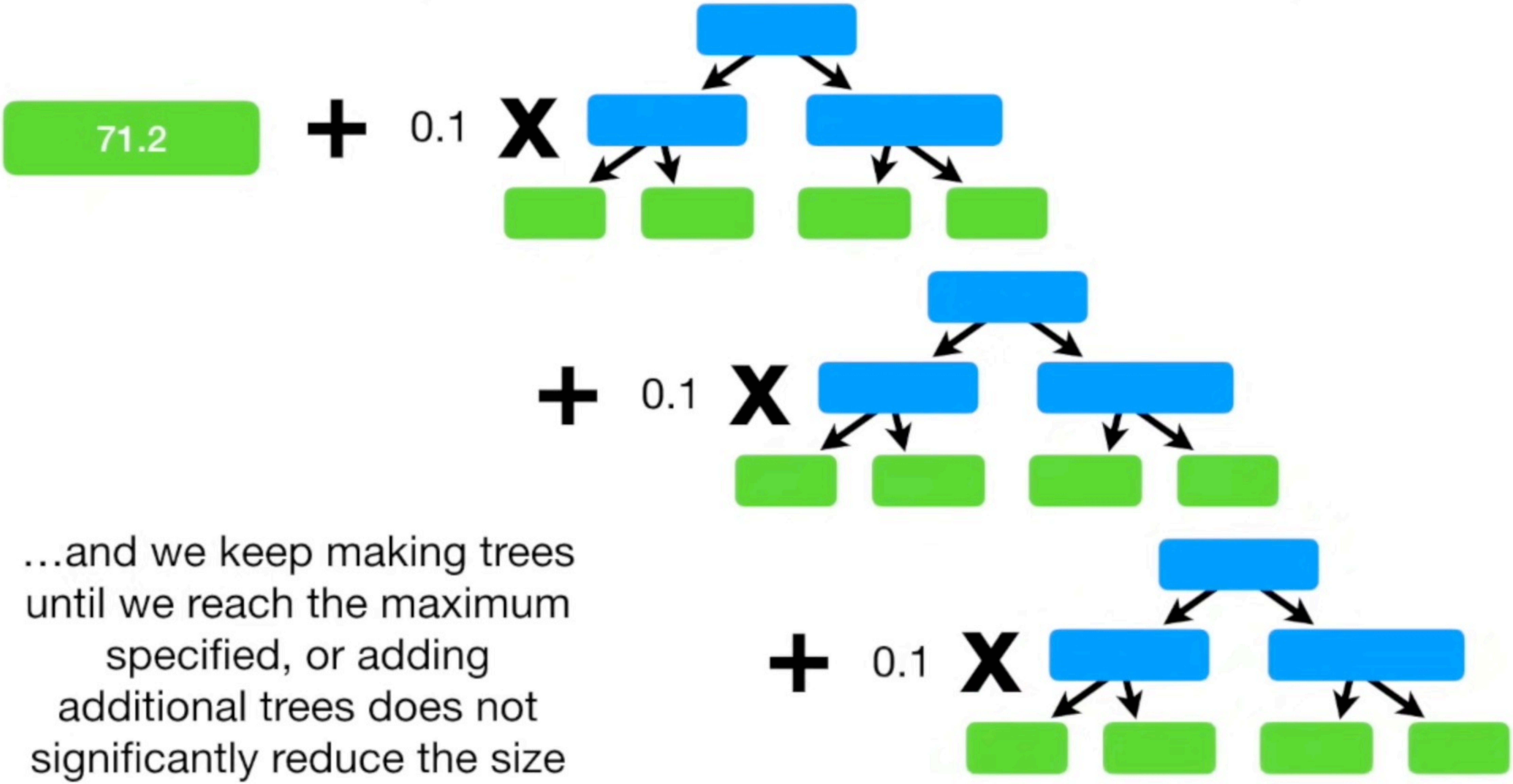
Residual	Residual	Residual
16.8	15.1	13.6
4.8	4.3	3.9
-15.2	-13.7	-12.4
1.8	1.4	1.1
5.8	5.4	5.1
-14.2	-12.7	-11.4

So we've taken another small step towards making good **Predictions**.



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

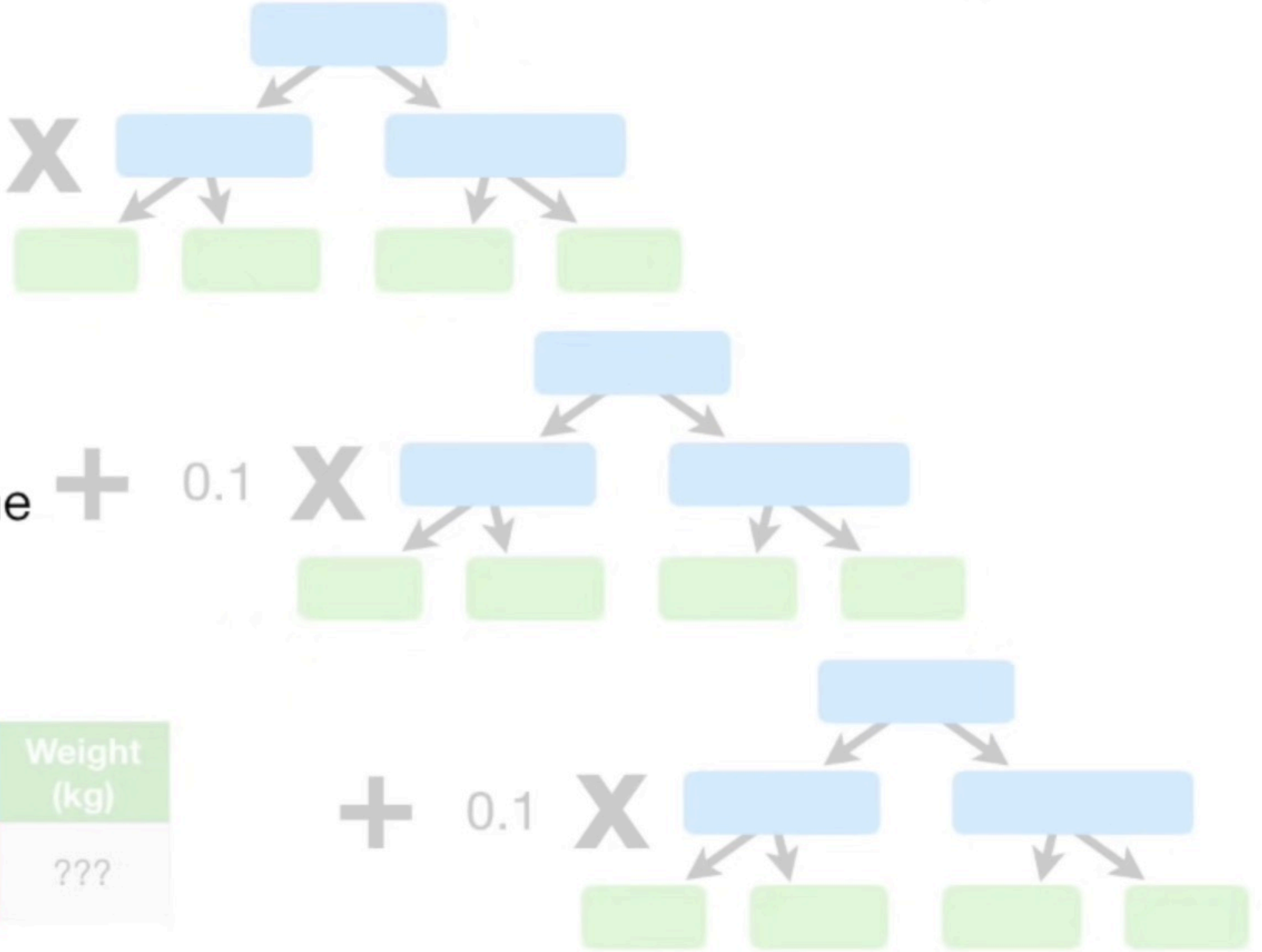
Now we build a another tree to predict the new **Residuals**…



...and 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...

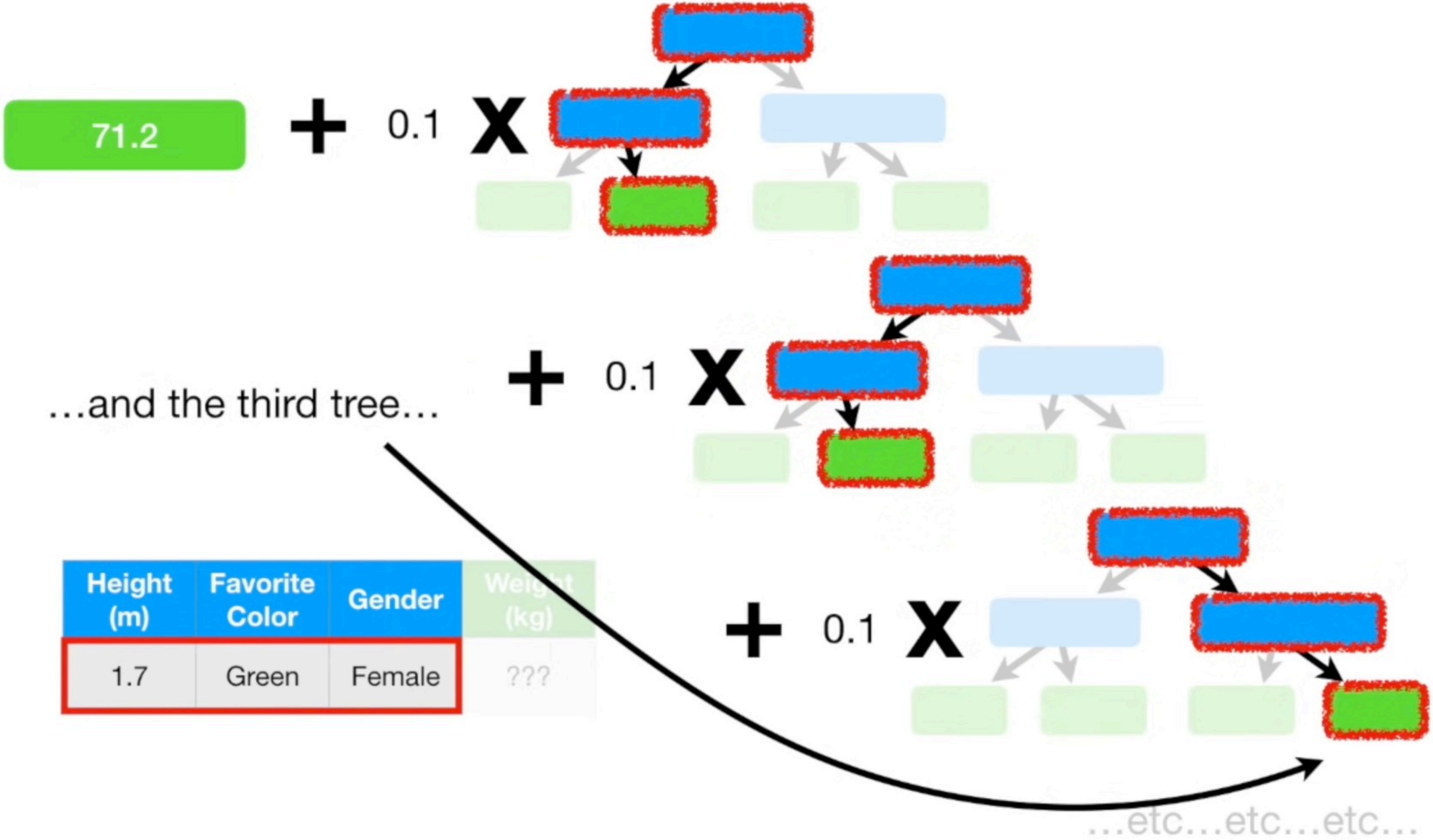
71.2 +



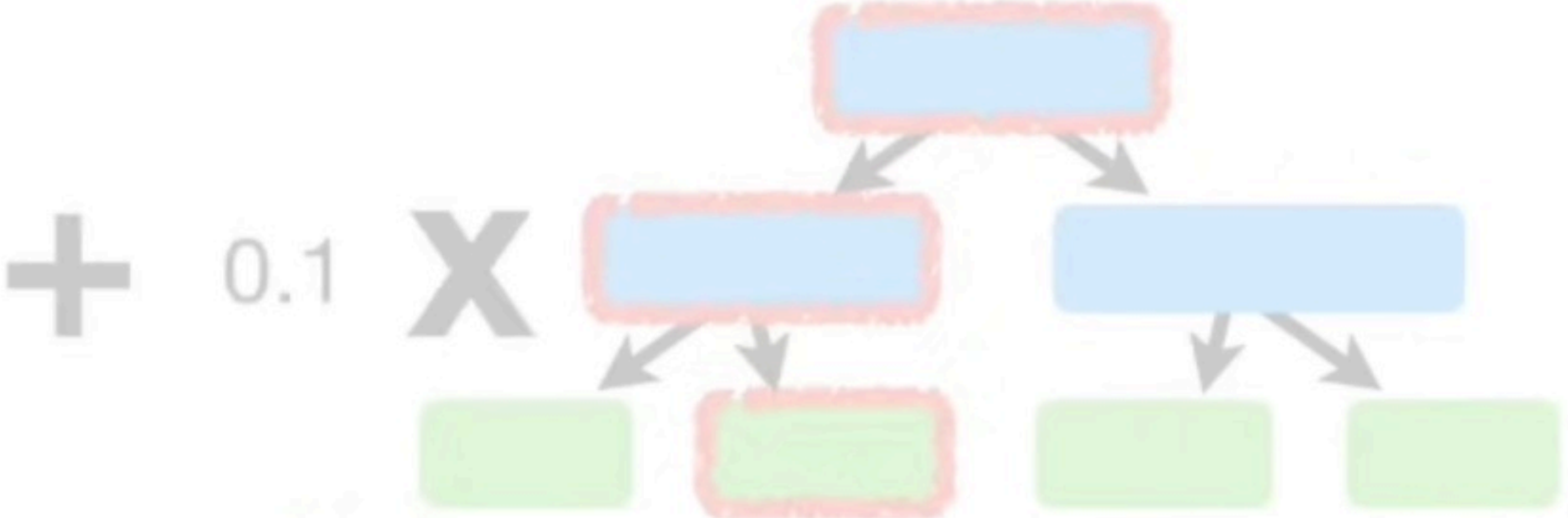
Then, when we get some new measurements…

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

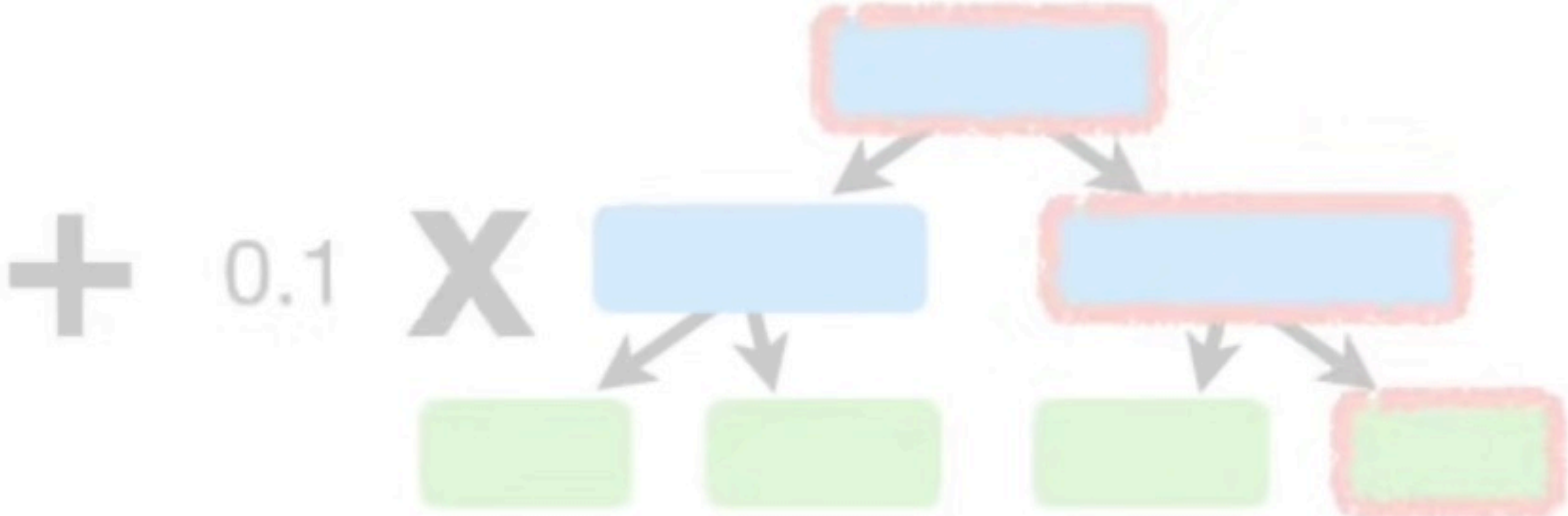
…etc…etc…etc…


71.2
+ 0.1 X
+ 0.1 X
+ 0.1 X
…and the third tree…
Height (m)
Favorite Color
Gender
Weight (kg)
1.7
Green
Female
???
…etc…etc…etc…

...and once the math is all done, we are left with the **Predicted Weight**.



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



...etc...etc...etc...