



Accuracy Paradox

Confusion Matrix



		\hat{y} (Predicted DV)	
		0	1
y (Actual DV)	0	9,700	150 !
	1	50 !	100

True positives

False negative

False Positives

True Negatives

		\hat{y} (Predicted DV)	
		0	1
y (Actual DV)	0	9,700	150 
	1	50 	100

Scenario 1:

Accuracy Rate = Correct / Total
 $AR = 9,800/10,000 = 98\%$

***This is a confusion matrix when we Are using a machine learning model.**

We can calculate the accuracy and the error Rates using the TP, TN, FP and FN.

		\hat{y} (Predicted DV)	
		0	1
y (Actual DV)	0	9,700 ← 150	150 ⚠
	1	50 ⚠ ← 100	100

Scenario 1:

Accuracy Rate = Correct / Total
 $AR = 9,800/10,000 = 98\%$

*Let's assume we decide to not use a model
 And hence we would predict everything as a 0



The new confusion matrix would look like this -

		\hat{y} (Predicted DV)	
		0	1
y (Actual DV)	0	9,850	0
	1	150	0

Scenario 1:

Accuracy Rate = Correct / Total
AR = 9,800/10,000 = 98%

The accuracy of scenario 1 is less than the accuracy of scenario 2, Whereas scenario 2 did not even use a model.

		\hat{y} (Predicted DV)	
		0	1
y (Actual DV)	0	9,850	0 
	1	150 	0

Scenario 1:

Accuracy Rate = Correct / Total
AR = 9,800/10,000 = 98%

Scenario 2:

Accuracy Rate = Correct / Total
AR = 9,850/10,000 = 98.5%

		\hat{y} (Predicted DV)	
		0	1
y (Actual DV)	0	9,850 ←	0 ⚠
	1	⚠ 150 ←	0

Scenario 1:

Accuracy Rate = Correct / Total
 AR = 9,800/10,000 = 98%

Scenario 2:

Accuracy Rate = Correct / Total
 AR = 9,850/10,000 = 98.5% ↑