

Matriz de Confusão



Acertos ou Acurácia

- Total que meu modelo acertou

Matriz de Treino			
Bom			500
Mau			500
Total			1000

Matriz Confusão			
		Bom	Mau
Original	Bom	450	50
	Mau	200	300

Os acertos estão nesta diagonal:

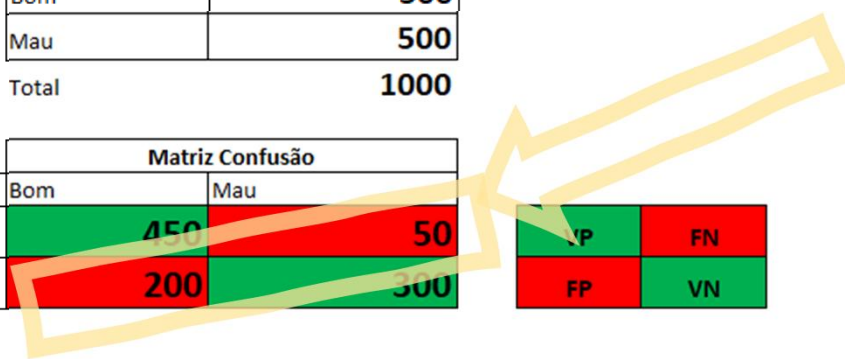
$$\frac{1000 \rightarrow 100}{750 \rightarrow X}$$
$$a = \frac{VP + VN}{T} = 0,75$$

Erros

- Total que meu modelo errou

Matriz de Treino			
Bom			500
Mau			500
Total			1000

Original	Matriz Confusão			
		Bom	Mau	
	Bom	450	50	FP FN
	Mau	200	300	
		VN VP		



Os erros estão nesta diagonal:

$$\frac{1000 \rightarrow 100}{250 \rightarrow X}$$

$$e = \frac{FP + FN}{T} = 0,25$$

Acertos ou Acurácia

Matriz de Treino	
Bom	900
Mau	100
Total	1000

		Matriz Confusão	
		Bom	Mau
Original	Bom	900	0
	Mau	100	0

VP	FN
FP	VN

Acurácia: 90%

Positivos Verdadeiros (Recall ou Sensivity)

- De todos os "Bons", quantos foram classificados corretamente como Bons?

Matriz de Treino			
Bom			500
Mau			500
Total			1000

Original	Matriz Confusão		
	Bom	Mau	
Bom	450	50	VP FN
Mau	200	300	FP VN

$$\frac{500 \rightarrow 100}{450 \rightarrow X}$$

$$pv = \frac{VP}{VP + FN} = 0,9$$

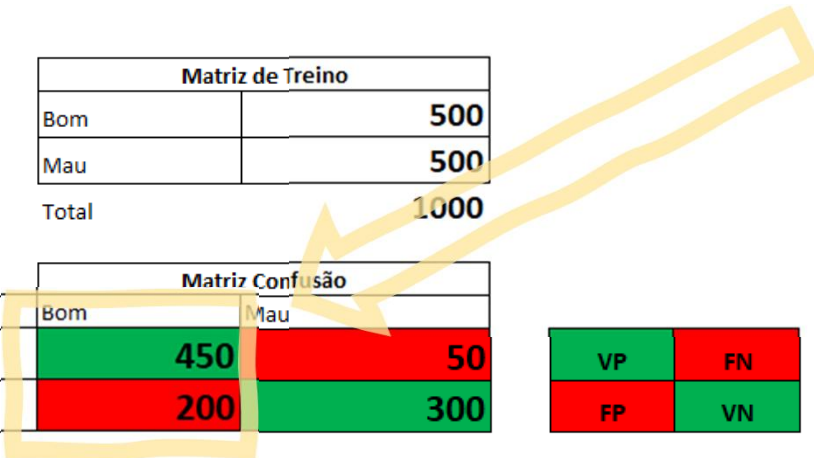
Precisão / Precision

- Dos que classifiquei como bons, quantos de fato eram bons?

Matriz de Treino			
Bom		500	
Mau		500	
Total		1000	

Matriz Confusão			
Original		Bom	Mau
	Bom	450	50
	Mau	200	300

VP	FN
FP	VN



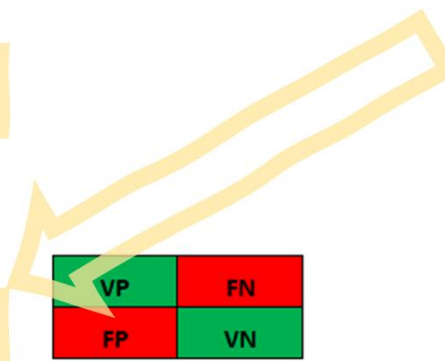
$$p = \frac{VP}{VP + FP} = 0,69$$

Negativos Verdadeiros: Specificity

- De todos os "maus", quantos foram classificados corretamente como "maus"?

Matriz de Treino	
Bom	500
Mau	500
Total	1000

		Matriz Confusão	
		Bom	Mau
Original	Bom	450	50
	Mau	200	300



VP	FN
FP	VN

$$nv = \frac{VN}{VN + FP} = 0,6$$

Positivos Falsos (Erro tipo I)

- De todos os "maus", quantos foram classificados incorretamente como "bons"?

Matriz de Treino			
Bom			500
Mau			500
Total			1000

Original	Matriz Confusão			
		Bom	Mau	
Original	Bom	450	50	VP
	Mau	200	300	FP

VP	FN
FP	VN

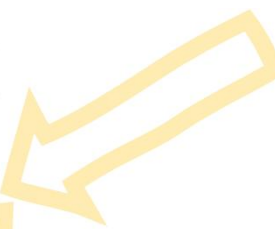
$$pf = \frac{FP}{VN + FP} = 0,4$$

Negativos Falsos (Erro Tipo II)

- De todos os "bons", quantos foram classificadas incorretamente como "maus"?

Matriz de Treino			
Bom			500
Mau			500
Total			1000

Matriz Confusão			
Original		Bom	Mau
	Bom	450	50
	Mau	200	300



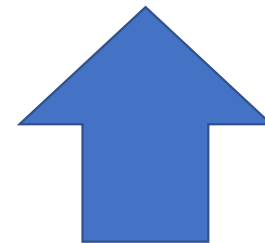
VP	FN
FP	VN

$$nf = \frac{FN}{VP + FN} = 0,1$$

F1 Score

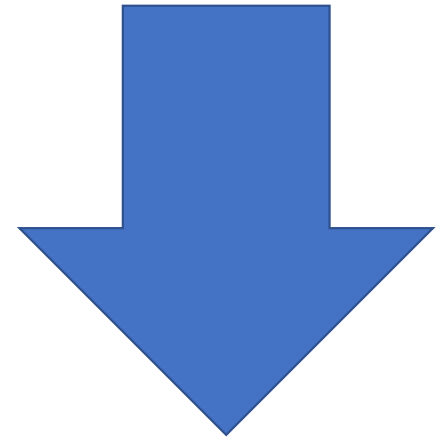
- É uma métrica de precisão
- Média harmônica entre Precision e Recall
- Valor entre zero e 1
- O objetivo é maximizar!
 - Cálculo, onde p é a precisão e r a lembrança

$$F1 = \frac{1}{\frac{1}{p} + \frac{1}{r}}$$

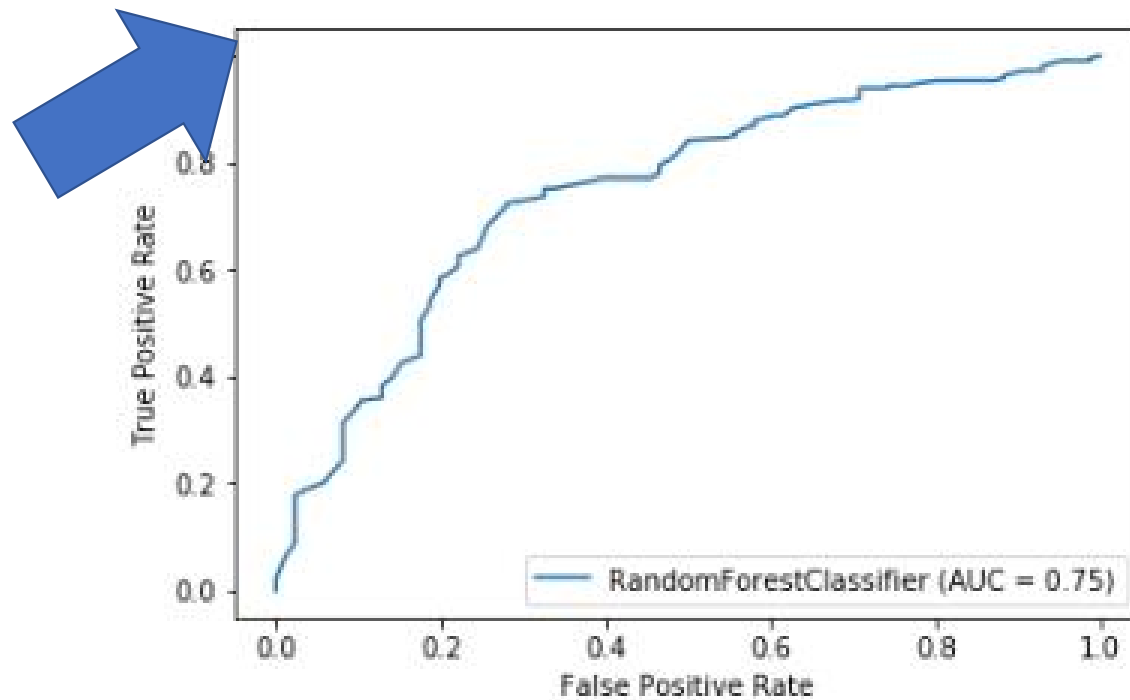


Logarithmic Loss

- Mostra a probabilidades para cada classe
- Bom para classificação multi-classe
- Penaliza a classificação errada
- Limite inferior zero
- Objetivo é minimizar



Receiver Operator Characteristic (ROC)



AUC: Area Under Curve

- É a área sobre a curva de Falsos Positivos versus Faltos Negativos
- Pode ter um valor entre zero e 1
- Quanto maior, melhor

