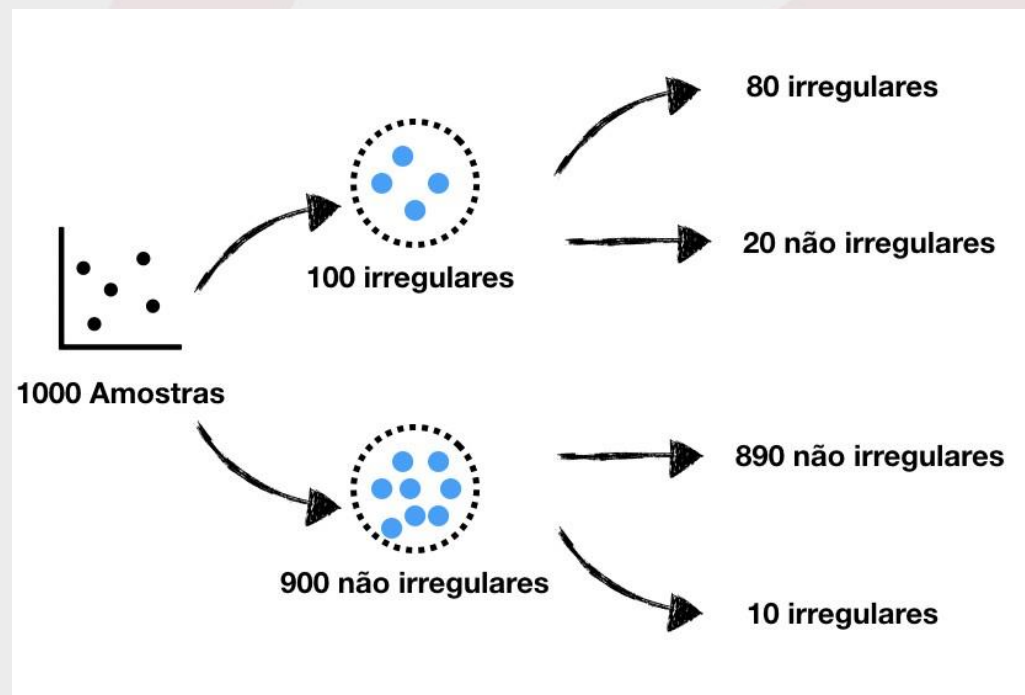


Aprendizado de Máquina

- Matriz de Confusão:
 - É uma tabela que permite a visualização do desempenho de um algoritmo de classificação.



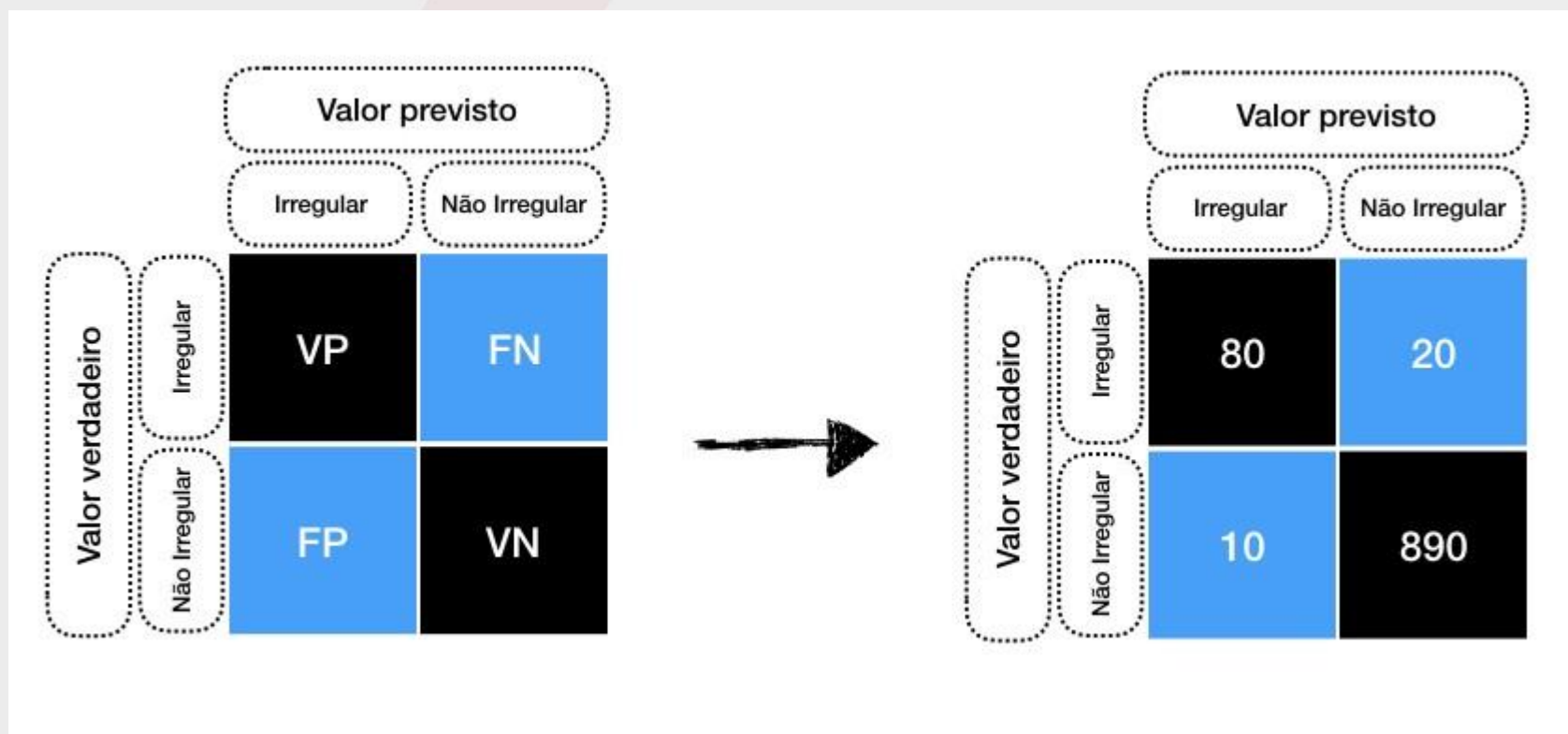
Aprendizado de Máquina

- Matriz de Confusão:
 - É uma tabela que permite a visualização do desempenho de um algoritmo de classificação.

		Verdade	
População total		Condição positiva	Condição negativa
Predito	Condição positiva prevista	Verdadeiro positivo	Falso positivo, Erro do tipo I
	Condição negativa prevista	Falso negativo, Erro do tipo II	Verdadeiro negativo

Aprendizado de Máquina

- Matriz de Confusão:



Aprendizado de Máquina

- Métricas:
 - **Verdadeiro positivo (VP)**: quando o método diz que a classe é positiva e, ao verificar a resposta, vê-se que a classe era realmente positiva;
 - **Verdadeiro negativo (VN)**: quando o método diz que a classe é negativa e, ao verificar a resposta, vê-se que a classe era realmente negativa;
 - **Falso positivo (FP)**: quando o método diz que a classe é positiva, mas ao verificar a resposta, vê-se que a classe era negativa;
 - **Falso negativo (FN)**: quando o método diz que a classe é negativa, mas ao verificar a resposta, vê-se que a classe era positiva.

Aprendizado de Máquina

- A acurácia avalia o percentual de acertos, ou seja, ela pode ser obtida pela razão entre a quantidade de acertos e o total de entradas:

$$acurácia = \frac{VP + VN}{VP + FN + VN + FP}$$

$$Acurácia = 80 + 890 / 80 + 20 + 10 + 890 = 970/1000 = 97,0\%$$

Aprendizado de Máquina

- Erro é igual a 1 - acurácia:

$$\text{Erro} = 1 - 0,97 = 3\%$$

Aprendizado de Máquina

- A precisão é uma métrica que indica, das classificações positivas do modelo, quantas foram acertadas.

Verdadeiros Positivos = 80, Falsos Positivos = 10

$$\text{Precisao} = \frac{\text{Verdadeiros Positivos}}{(\text{Verdadeiros Positivos} + \text{Falsos Positivos})}$$

$$\text{Precisao} = \frac{80}{(80 + 10)}$$

$$\text{Precisao} = \frac{80}{90}$$

$$\text{Precisao} = 0.8888 \text{ (x 100)} = 88.88 \%$$

Aprendizado de Máquina

- A revocação (sensibilidade ou *recall*) é uma métrica que indica, das amostras positivas existentes, quantas o modelo conseguiu classificar corretamente.

Verdadeiros Positivos = 80, Falsos Negativos = 20

$$\text{Revocacao} = \frac{\text{Verdadeiros Positivos}}{(\text{Verdadeiros Positivos} + \text{Falsos Negativos})}$$

$$\text{Revocacao} = \frac{80}{(80 + 20)}$$

$$\text{Revocacao} = \frac{80}{100}$$

$$\text{Revocacao} = 0.8 (x 100) = 80 \%$$

Aprendizado de Máquina

- A especificidade avalia a capacidade do método de detectar resultados negativos.

$$especificidade = \frac{VN}{VN + FP}$$

$$\text{Especificidade} = 890 / 890 + 10 = 890/900 = 98,9\%$$

Aprendizado de Máquina

- F-measure, F-score ou score F1 é uma média harmônica calculada com base na precisão e na revocação.

$$f1 = 2 * \frac{\textit{precisão} * \textit{sensibilidade}}{\textit{precisão} + \textit{sensibilidade}}$$

$$F1 = 2 * 0,88 * 0,80 / 0,88 + 0,80 = 2 * 0,704 / 1,68 = 83,8\%$$

Aprendizado de Máquina

- Navalha de Occam
 - Do latim “lex parsimoniae” – A lei da parcimônia (economia)
 - A explicação de qualquer fenômeno deve fazer o menor número possível de suposições, eliminando aquelas que não fazem diferença nas previsões observáveis da hipótese explicativa ou da teoria.
 - O Dilema de Occam: Infelizmente, em AM, acurácia e simplicidade estão em conflito.

Aprendizado de Máquina

- Maldição da dimensionalidade:
 - Aprender a partir de um espaço de característica de alta dimensionalidade requer uma quantidade enorme de dados de treinamento para garantir que haja várias amostras com cada combinação de valores.
 - Com uma quantidade fixa de número de instâncias de treinamento, o poder de preditibilidade reduz à medida que aumenta a dimensionalidade.

