



# **Deteccção de Fraudes**



# **Machine Learning**

## SUPERVISIONADO

**UM MODELO PARA  
AJUSTAR**

**UMA MÉTRICA PARA  
OTIMIZAR**

**UM MÉTODO PARA  
VALIDAR**



# **Amostra Estratificada**



# **Métricas**

## Classificação Binária

# Métricas

## Classificação Binária

	Preditos	Observados
0	1	1
1	0	0
2	0	0
3	1	0
4	0	0
...	...	...
1333	0	0
1334	0	0
1335	0	0
1336	0	0
1337	1	1

Matriz de Confusão		RESPOSTA PREDITA	
		0	1
RESPOSTA OBSERVADA	0	VERDADEIRO NEGATIVO (VN)	FALSO POSITIVO (FP)
	1	FALSO NEGATIVO (FN)	VERDADEIRO POSITIVO (VP)


$$\text{Sensibilidade} = \frac{VP}{VP + FN}$$


Também chamada de **recall, taxa de verdadeiro positivos.**

		PREDITO	
		0	1
OBSERVADO	0	VN	FP
	1	FN	VP


$$\text{Valor Preditivo Positivo} = \frac{VP}{VP + FP}$$

Também chamada de **precisão**.

		PREDITO	
		0	1
OBSERVADO	0	VN	FP
	1	FN	VP


$$\text{F1-Score} = \frac{2 * \text{Sensibilidade} * \text{VPP}}{\text{Sensibilidade} + \text{VPP}}$$

É a média harmônica entre  
**Sensibilidade e Valor  
Preditivo Positivo.**

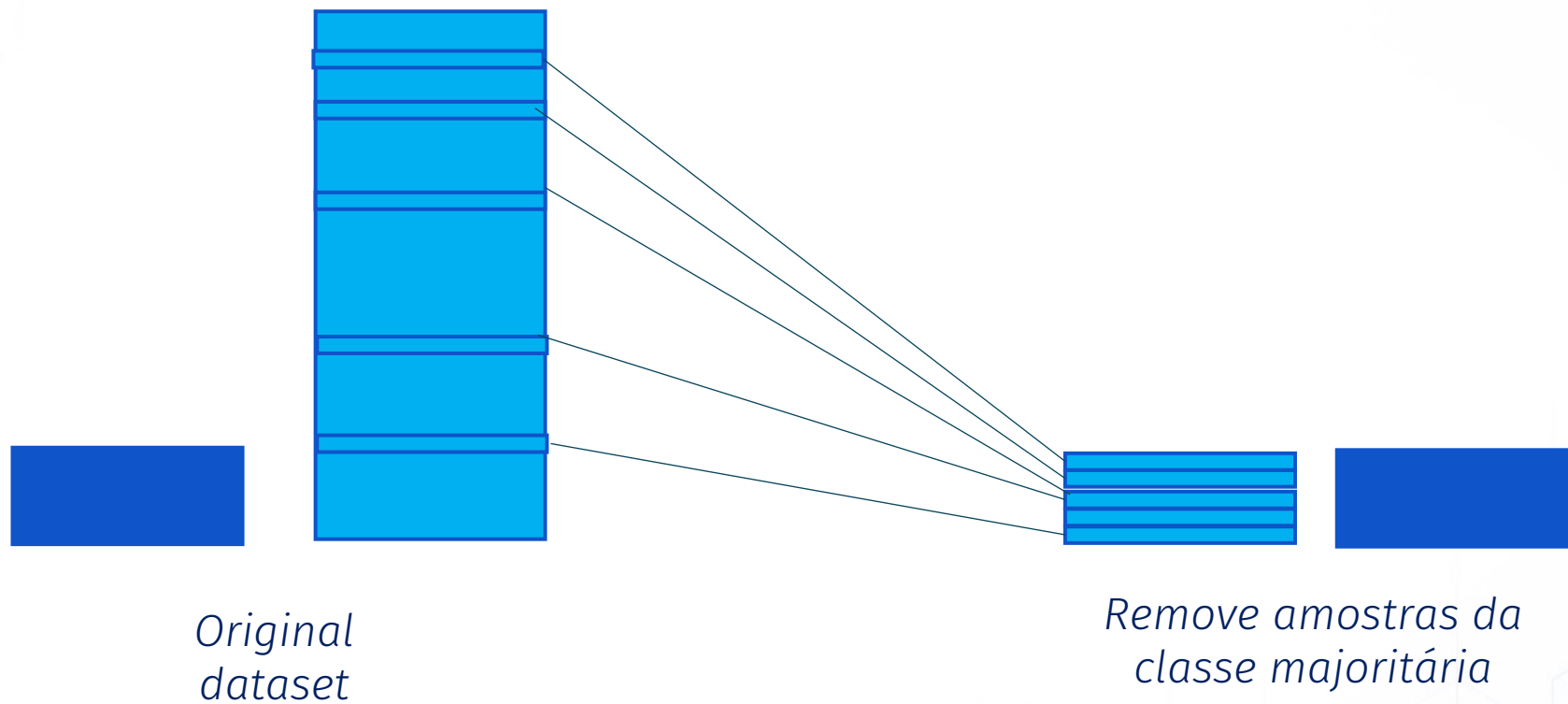
		PREDITO	
		0	1
OBSERVADO	0	VN	FP
	1	FN	VP





# **Dados Desbalanceados**

# Undersampling



# Oversampling

