

# Concessão de Crédito

---

ID	Age	Income	CreditScore	LoanAmount	PaymHistory	Class	Prediction	PredictionAccuracy
1	35	75000	720	25000	0	good	good	Correto
2	28	48000	650	15000	2	bad	bad	Correto
3	42	85000	710	35000	1	good	good	Correto
4	52	120000	680	50000	3	bad	good	Incorreto
5	38	60000	640	15000	0	good	good	Correto
6	29	50000	590	10000	4	bad	bad	Correto
7	47	92000	750	45000	1	good	bad	Incorreto
8	31	55000	600	20000	3	bad	good	Incorreto
9	45	100000	730	30000	0	good	good	Correto
10	39	67000	625	18000	2	bad	bad	Correto

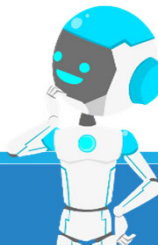


# Matriz de Confusão (Classificação)

ID	Class	Prediction	PredictionAccuracy	
1	good	good	Correto	VP
2	bad	bad	Correto	VN
3	good	good	Correto	VP
4	bad	good	Incorreto	FP
5	good	good	Correto	VP
6	bad	bad	Correto	VN
7	good	bad	Incorreto	FN
8	bad	good	Incorreto	FP
9	good	good	Correto	VP
10	bad	bad	Correto	VN

	Classe Real Positiva	Classe Real Negativa
Classe Prevista Positiva	Verdadeiros Positivos	Falsos Negativos
Classe Prevista Negativa	Falsos Positivos	Verdadeiros Negativos

	Verd. Good	Verd. Bad
Prev Good	4	1
Prev Bad	2	3



# Acurácia

---

	Classe Real Positiva	Classe Real Negativa
Classe Prevista Positiva	Verdadeiros Positivos	Falsos Negativos
Classe Prevista Negativa	Falsos Positivos	Verdadeiros Negativos

## Classificações corretas

Formula:  $(VP + VN) / (VP + VN + FP + FN)$

Calculation:  $(4 + 3) / (4 + 3 + 2 + 1) = 7 / 10 = 0.7$

Interpretação: O modelo classificou corretamente 70% das instâncias

	Verd. Good	Verd. Bad
Prev Good	4	1
Prev Bad	2	3



# Precisão

A proporção de instâncias verdadeiramente positivas entre as instâncias previstas como positivas.

**Fórmula:**  $VP / (VP + FP)$

**Cálculo:**  $4 / (4 + 2) = 4 / 6 = 0,67$

**Interpretação:** Das instâncias previstas como positivas, 67% eram realmente positivas.

ID	Class	Prediction	PredictionAccuracy	
1	good	good	Correto	VP
2	bad	bad	Correto	VN
3	good	good	Correto	VP
4	bad	good	Incorreto	FP
5	good	good	Correto	VP
6	bad	bad	Correto	VN
7	good	bad	Incorreto	FN
8	bad	good	Incorreto	FP
9	good	good	Correto	VP
10	bad	bad	Correto	VN

	Classe Real Positiva	Classe Real Negativa
Classe Prevista Positiva	Verdadeiros Positivos	Falsos Negativos
Classe Prevista Negativa	Falsos Positivos	Verdadeiros Negativos

	Verd. Good	Verd. Bad
Prev Good	4	1
Prev Bad	2	3



# Recall

Recall (Sensibilidade ou Taxa de Verdadeiros Positivos): A proporção de instâncias classificadas como positivas em relação às instâncias positivas reais.

**Fórmula:**  $VP / (VP + FN)$

**Cálculo:**  $4 / (4 + 1) = 4 / 5 = 0,8$

**Interpretação:** O modelo identificou corretamente 80% das instâncias boas.

ID	Class	Prediction	PredictionAccuracy	
1	good	good	Correto	VP
2	bad	bad	Correto	VN
3	good	good	Correto	VP
4	bad	good	Incorreto	FP
5	good	good	Correto	VP
6	bad	bad	Correto	VN
7	good	bad	Incorreto	FN
8	bad	good	Incorreto	FP
9	good	good	Correto	VP
10	bad	bad	Correto	VN

	Classe Real Positiva	Classe Real Negativa
Classe Prevista Positiva	Verdadeiros Positivos	Falsos Negativos
Classe Prevista Negativa	Falsos Positivos	Verdadeiros Negativos

	Verd. Good	Verd. Bad
Prev Good	4	1
Prev Bad	2	3



# Especificidade

**Especificidade:** A proporção de instâncias negativas verdadeiras das instâncias negativas reais.

**Fórmula:**  $VN / (VN + FP)$

**Cálculos:**  $3 / (3 + 1) = 3 / 4 = 0.75$

**Interpretação:** Identificou corretamente 75% das instâncias negativas como negativas

ID	Class	Prediction	PredictionAccuracy	
1	good	good	Correto	VP
2	bad	bad	Correto	VN
3	good	good	Correto	VP
4	bad	good	Incorreto	FP
5	good	good	Correto	VP
6	bad	bad	Correto	VN
7	good	bad	Incorreto	FN
8	bad	good	Incorreto	FP
9	good	good	Correto	VP
10	bad	bad	Correto	VN

	Classe Real Positiva	Classe Real Negativa
Classe Prevista Positiva	Verdadeiros Positivos	Falsos Negativos
Classe Prevista Negativa	Falsos Positivos	Verdadeiros Negativos

	Verd. Good	Verd. Bad
Prev Good	4	2
Prev Bad	1	3



# F1 Score

**F1 Score:** A média harmônica de precisão e recall

**Formula:**  $2 * (\text{Precisão} * \text{Recall}) / (\text{Precisão} + \text{Recall})$

**Cálculo:**  $2 * (0.67 * 0.8) / (0.67 + 0.8) = 2 * (0.536) / 1.47 = 0.73$

**Interpretação:** 0.73 Indica que há um equilíbrio entre precisão e recall

	Classe Real Positiva	Classe Real Negativa
Classe Prevista Positiva	Verdadeiros Positivos	Falsos Negativos
Classe Prevista Negativa	Falsos Positivos	Verdadeiros Negativos

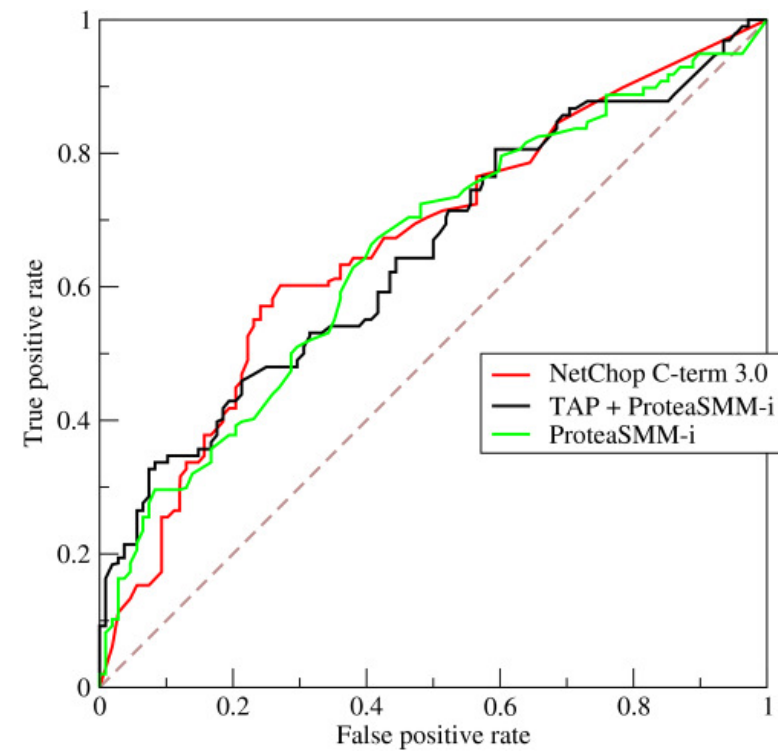
	Verd. Good	Verd. Bad
Prev Good	4	2
Prev Bad	1	3



# ROC

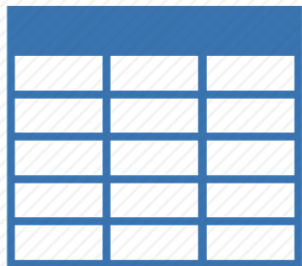
O gráfico ROC (Receiver Operating Characteristic ) e a métrica Area Under the ROC Curve (AUC-ROC) são usados para avaliar a performance de classificação binária.

O AUC-ROC pode variar entre 0 e 1. Quanto mais próximo de 1, melhor a performance do classificador.





# Problema da Classe Rara







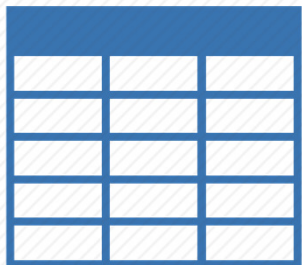
1 milhões de transações  
5% fraudulentas



Amostragem Aleatória  
sem reposição: 1%



10.000  
9.900 - Legítimas  
100 - Fraudulentas







1 milhões de transações  
5% fraudulentas



Amostragem Aleatória  
sem reposição:  
0,5% - Legítimas  
0,5% - Fraudulentas



10.000  
5.000 - Legítimas  
5.000 - Fraudulentas



# Não devemos olhar apenas acurácia!

## Caso 1: Dataset desbalanceado

- 990 bad
- 10 good

Modelo sempre classifica tudo como bad

- Acurácia: 99%
- Recall: 0%

	Classe Real Positiva	Classe Real Negativa
Classe Prevista Positiva	Verdadeiros Positivos	Falsos Negativos
Classe Prevista Negativa	Falsos Positivos	Verdadeiros Negativos

	Verd. Good	Verd. Bad
Prev Good	0	0
Prev Bad	10	900



# Não devemos olhar apenas acurácia!

## Caso 1: Custos Diferentes

### Doença Rara

- 950 Saudável
- 50 Doente

### Métricas

- Acurácia: 96%
- Precisão: 67%
- Recall: 40%
- F1 Score: 50%

	Classe Real Positiva	Classe Real Negativa
Classe Prevista Positiva	Verdadeiros Positivos	Falsos Negativos
Classe Prevista Negativa	Falsos Positivos	Verdadeiros Negativos

	Verd. Saudável	Verd. Doente
Prev Saudável	950	30
Prev Doente	10	20



## Problemas de Atributos Desconhecidos

---

No treino: regiões  
“Sul”, “Sudeste”,  
“Centro-Oeste” e  
“Norte”

Na produção:  
região “Nordeste”

