

Aula 20 - Ferramentas para Construção de Pipelines de IA. Ajuste e Otimização de Modelos de IA. Balanceamento e Métricas.

Profa. Gabrielly Queiroz

Pipelines

Um pipeline é uma sequência automática de etapas que transforma os dados e treina o modelo, como uma linha de montagem.

https://colab.research.google.com/drive/1xSKRIG3xpRkARodOMd4S9IEF7XRRK_KIM?usp=sharing

Matriz Confusão

CLASSIFICAÇÃO DO MODELO

REAL

		acertos
		erros
SITUAÇÃO REAL	VERDADEIRO POSITIVO	VP 70
	VERDADEIRO NEGATIVO	VN 50
VERDADEIRO POSITIVO	FALSO POSITIVO	FP 30
	FALSO NEGATIVO	FN 10

VP - Verdadeiros Positivos
VN - Verdadeiros Negativos
FP - Falsos Positivos
FN - Falsos Negativos

Situação real	Previsão do modelo	Res.	Nome técnico
Pessoa tem a doença	Modelo diz que tem	A	Verdadeiro Positivo (TP)
Pessoa não tem a doença	Modelo diz que não tem	A	Verdadeiro Negativo (TN)
Pessoa não tem a doença	Modelo diz que tem	E	Falso Positivo (FP)
Pessoa tem a doença	Modelo diz que não tem	E	Falso Negativo (FN)

Métricas

Precision (Precisão):

- Mede **quantas das previsões positivas estavam corretas.**

Alta precisão → poucos falsos positivos.

Recall (Revocação ou Sensibilidade):

- Mede **quantas das amostras positivas foram detectadas.**

Alto recall → poucos falsos negativos.

F1-Score:

- Média harmônica entre *precision* e *recall*.
- Equilibra os dois indicadores.

Ideal quando há **dados desbalanceados**.

Accuracy (Acurácia):

- Percentual de acertos totais (todas as classes).

Pode ser **enganosa em bases desbalanceadas**.

$$\text{Precision} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Positivos}}$$

$$\text{Recall} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Negativos}}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

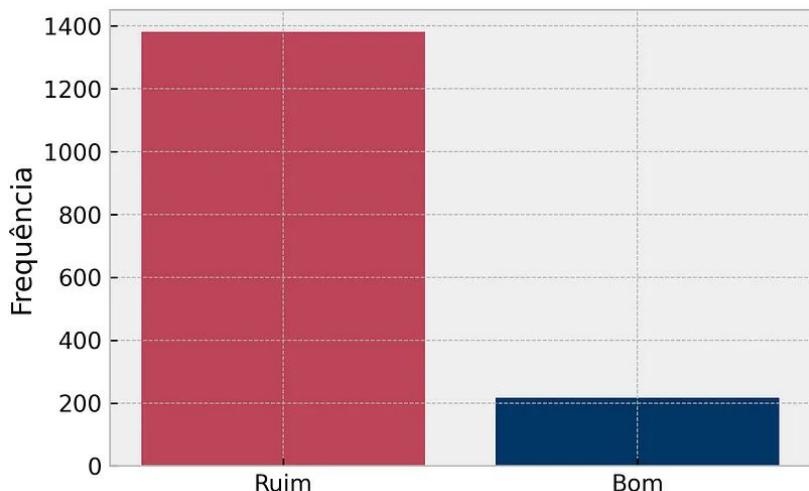
$$\text{Accuracy} = \frac{\text{Acertos Totais}}{\text{Total de Amostras}}$$

Balanceamento dos Dados

O modelo aprende a **favorecer a classe majoritária**, resultando em alta acurácia, mas **baixa detecção da classe rara**.

Problemas comuns:

- Falsos negativos altos (ignora eventos importantes).
- Acurácia enganosa.
- Métricas tradicionais (accuracy) se tornam pouco informativas.



Balanceamento dos Dados

SMOTE — Synthetic Minority Over-sampling Technique:

- Cria **novas amostras sintéticas** da classe minoritária, em vez de apenas duplicar exemplos existentes.
- Gera pontos *intermediários* entre exemplos reais próximos.

`class_weight='balanced'` (do Scikit-learn)

- **Não altera os dados.**
- Ajusta **os pesos de cada classe** dentro do cálculo do erro.
- O modelo **pune mais** os erros da classe minoritária.

