WORKSHOP DE ESTUDANTES 2

COMO SABEMOS SE O MODELO É BOM? TREINO E TESTE, MATRIZ DE CONFUSÃO, NOTAS, SUBAJUSTE E SOBREAJUSTE

Sistema de Monitorização de Recursos Hídricos · Rio Incomáti, Marracuene

Amilton Docole & Agostinho Oliveira

Sábado, 21 de Novembro de 2026 · 08h–14h · Maputo

Como é que um professor sabe se um aluno aprendeu a matéria, ou se só decorou as respostas?

O CAMINHO COMPLETO DO TREINO

Passo 1
Pergunta
O que queremos prever?

Passo 2
Dados
Imagens, séries, tabelas.

Passo 3
Rótulos
As respostas certas (só na supervisionada).

Passo 4
Características
As pistas de cada exemplo.

Passo 5
Dividir
Treino de um lado, teste do outro.

Passo 6
Treinar
modelo.fit(X, y)

Passo 7
Avaliar
As notas no teste.

Passo 8
Usar
modelo.predict(X_novo)

Se a nota do teste for má, volta-se atrás: mais exemplos, outras pistas, outro algoritmo.

TREINO E TESTE: ESTUDAR E FAZER O EXAME

Treino, cerca de 70%: o modelo aprende

Teste, cerca de 30%: fechado à chave

Treino = estudar
Os exemplos com rótulo. O modelo aprende com eles.
modelo.fit(X, y)

Teste = o exame
Dados que o modelo NUNCA viu. Servem só para dar a nota.
modelo.predict(X_teste)

QUANDO O TEMPO IMPORTA, O EXAME É O FUTURO

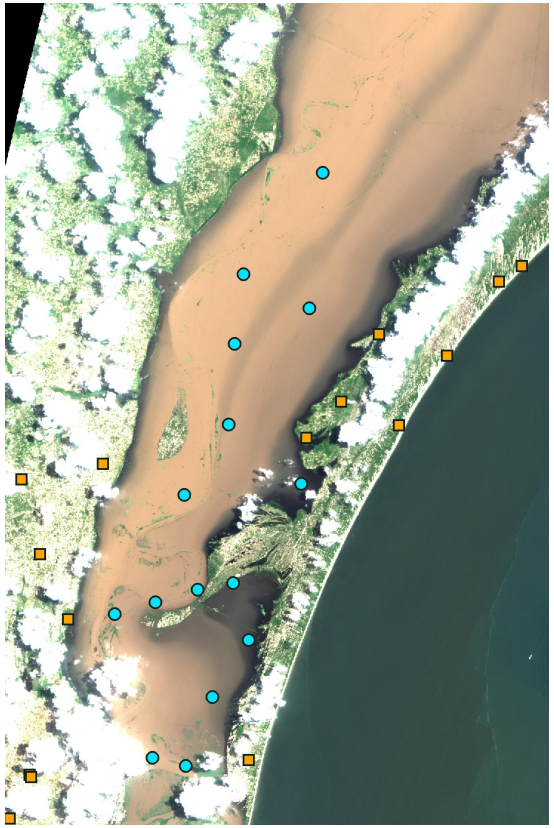
Treino: 2006 2018

Exame: 2019–2026

Na previsão do caudal (Parte F), o modelo aprende com o passado e é avaliado com anos que nunca viu, incluindo a cheia de 2026.

Misturar os dias ao acaso seria copiar: um dia e o seguinte são quase iguais.

NO NOSSO RIO: OS RÓTULOS SÃO PONTOS MARCADOS



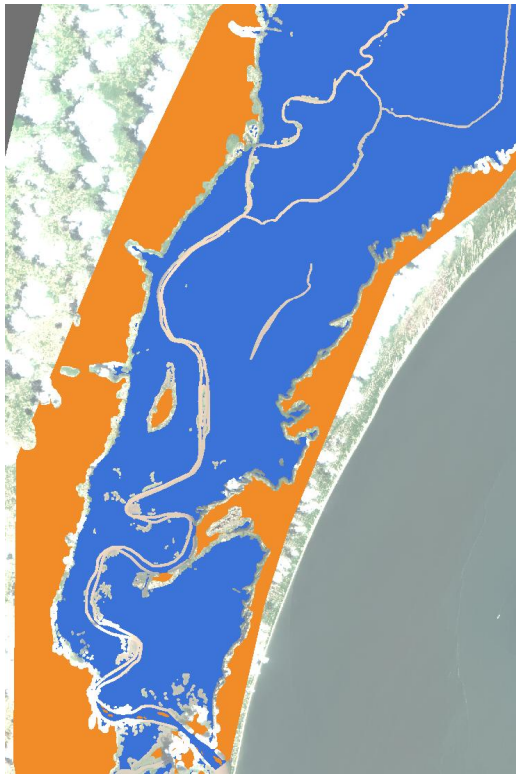
Uma pessoa olha para a imagem e marca pontos: **círculos azuis = água**, **quadrados laranja = não água**. Cada ponto vale 25 píxeis à sua volta.

Bons Rótulos são variados
Água barrenta, rio, lagoa escura. Mato, campos, areia, vila. Longe das nuvens e das bordas.

Rótulos errados ensinam coisas erradas.

Figura 1. Os 15 + 15 pontos da equipa na imagem da cheia (24/01/2026). Fonte: Copernicus Sentinel-2 (ESA).

REGRA DE OURO: UM EXAME INDEPENDENTE



Nunca dar a nota com rótulos feitos pelo próprio modelo.

O nosso exame: as amostras da cheia de janeiro de 2026 do Copernicus EMS, o serviço europeu de mapas de emergência. O modelo nunca as viu.

Azul = água; laranja = não água

Figura 2. As amostras do exame sobre a imagem da cheia. Fonte: Copernicus EMS (EMSR857) e JRC Global Surface Water.

REGRA DE OURO: UM EXAME INDEPENDENTE

Até o exame tinha um erro

O problema

O EMS só marca a água da cheia, não o rio de sempre: algumas amostras de “não água” estavam em cima do rio.

A correcção

Juntámos 38 anos de água vista pelo Landsat (arquivo JRC): o rio de sempre passou a contar como água.

Um exame também pode errar. Quando o modelo e o exame discordam, vale a pena ir ver quem tem razão.