

INSIGHTS · TUTORIAL DA SEMANA

Detetar anomalias em séries temporais com Isolation Forest — método prático para reduzir falsos positivos

Desliza →

Por que detetar anomalias?

Manter qualidade de dados, monitorizar sistemas e avisar eventos raros em produção. Útil quando há poucas etiquetas de anomalia.

Pré-requisitos

Python 3.7+, pandas e scikit-learn; ficheiro CSV com timestamp e value. Recomendado 1k–100k pontos; séries muito curtas pedem validação extra.

Conhecer a frequência

Verifique se a série é horária, diária ou irregular. Reamostrar para frequência fixa antes de extrair características, se necessário.

Criar características temporais

Use janelas móveis, lags e variáveis cíclicas (hora, dia da semana). Exemplos de janelas: 3, 24, 168 (horária) ou 7, 30, 90 (diária).

Treinar Isolation Forest

Escale os dados e treine o modelo. Ajuste contamination conforme a prevalência esperada (ex.: $0.01 = 1\%$). `n_estimators=100` é um bom início.

Reduzir falsos positivos

Exigir persistência (por exemplo, 2+ pontos consecutivos), usar thresholds por percentil ou combinar com regras operacionais para filtrar alarmes.

Verificar e iterar

Inspeção amostras manuais (50–200 casos),
calcule métricas se tiver etiquetas e defina KPIs
de aceitabilidade (ex.: falsos positivos/dia).

CONTINUA NO BLOG

E na vossa realidade, como detetam e validam anomalias em séries temporais?

[**www.bconcepts.pt/insights**](http://www.bconcepts.pt/insights)