

# Detecção de Anomalias com Isolation Forest

O algoritmo **Isolation Forest** é uma técnica baseada em árvores de decisão projetada para identificar anomalias em conjuntos de dados multidimensionais. Diferente de métodos estatísticos tradicionais, ele isola as observações em vez de modelar regiões normais dos dados. O princípio é que anomalias são mais fáceis de isolar, exigindo menos divisões (ou partições) nas árvores.

## 1. Conceito do Algoritmo

O Isolation Forest constrói um conjunto de árvores de isolamento (ou subamostras aleatórias dos dados) onde cada nó divide aleatoriamente o espaço das variáveis. A ideia central é que observações anômalas são isoladas rapidamente, ou seja, possuem menores profundidades médias nas árvores.

## 2. Aplicação aos Dados de Condensação e Evaporação

No conjunto de dados utilizado, o modelo foi aplicado sobre as variáveis: '**condensacao**', '**evaporacao**' e '**superaquecimento**'. Essas variáveis representam medições físicas relevantes do sistema de refrigeração analisado.

O objetivo foi detectar observações fora do comportamento normal do sistema, como possíveis falhas ou medições inconsistentes. Para isso, foi treinado o modelo com os parâmetros definidos em um dicionário, posteriormente passado ao construtor do IsolationForest.

## 3. Parâmetros Utilizados

- 1 **n\_estimators**: 200 – número de árvores no ensemble.
- 2 **max\_samples**: 256 – número de amostras por árvore.
- 3 **contamination**: 0.05 – proporção estimada de anomalias.
- 4 **random\_state**: 42 – garante reprodutibilidade.

## 4. Interpretação dos Resultados

Após o ajuste, o método **predict()** retorna: **-1** para observações consideradas anômalas. **1** para observações normais. Já o método **decision\_function()** fornece o escore de anomalia, onde valores menores indicam maior probabilidade de serem pontos fora do padrão.

## 5. Visualização

Os gráficos de dispersão (*jointplots*) entre '**condensacao**' e '**evaporacao**' foram utilizados para inspecionar visualmente as anomalias, destacando em vermelho os pontos anômalos e em preto os pontos normais. Essa visualização auxilia na identificação de padrões e na verificação qualitativa do comportamento detectado pelo modelo.

## 6. Conclusão

O Isolation Forest demonstrou-se uma abordagem eficiente para identificar anomalias em sistemas térmicos e mecânicos, como o analisado, ao lidar bem com dados contínuos e não requerer suposições sobre a distribuição dos dados. A interpretação dos escores e a visualização conjunta reforçam a utilidade do método em ambientes industriais.