

# Documentação - Árvore de decisão

Uma **árvore de decisão** é um modelo de machine learning usado para tomar decisões com base em perguntas sequenciais. Ela funciona como um fluxograma, onde cada nó interno representa uma decisão (ou condição) e cada folha indica um resultado.

## Quando usar?

- Quando você precisa de **explicações claras** sobre como as decisões foram tomadas (como em diagnósticos médicos).
- Ideal para **classificação** (ex: "Este cliente vai cancelar a assinatura?") ou **regressão** (ex: "Qual será o valor de venda de um imóvel?").

## Vantagens

- **Fácil de interpretar:** Dá para entender a lógica olhando a estrutura da árvore.
- **Funciona com dados mistos:** Aceita dados numéricos e categóricos.
- **Rápida para treinar:** Principalmente com pequenas bases de dados.

## Desvantagens

- **Sensível a overfitting:** Pode se ajustar demais aos dados de treino e ter baixo desempenho em novos dados.
- **Menos precisa** que outros modelos como Random Forest ou Redes Neurais em casos mais complexos.

Quando uma árvore de decisão é treinada, ela precisa decidir **qual a melhor pergunta** para separar os dados. A ideia é fazer as divisões que organizem os dados da maneira mais clara possível. Para isso, usamos **métricas**. As principais são:

## Entropia

### O que é?

- É uma medida que mostra **o quão desordenado ou incerto** um conjunto de dados está. Quanto maior a entropia, mais "bagunçados" os dados estão.

### Como funciona?

Se os dados estiverem bem misturados (por exemplo, metade "Sim" e metade "Não"), a entropia é alta.

Se os dados estiverem organizados (por exemplo, 90% "Sim" e 10% "Não"), a entropia é baixa.

### **Quando usar?**

Usamos entropia quando queremos maximizar a **organização** após uma divisão.

## ***Gini (Índice Gini)***

### ***O que é?***

- O Gini mede **quão puros os grupos ficam** depois de uma divisão. Um grupo é "puro" quando quase todos os elementos são da mesma classe (por exemplo, todos são "Sim").

### ***Como funciona?***

Quanto mais puro o grupo, mais próximo de 0 o índice Gini será. Se a divisão for ruim, o Gini estará mais próximo de 0,5.

### ***Quando usar?***

O Gini é rápido e direto, sendo uma alternativa comum à entropia.

## ***Erro Quadrático Médio (MSE)***

- **O que é?**

Usado em problemas de **regressão** para medir **o quão distante a previsão está** do valor real.

### **Como funciona?**

Ele calcula a média das diferenças ao quadrado entre o valor real e o previsto. Quanto menor o MSE, melhor a previsão.

### **Quando usar?**

Ideal para prever números contínuos, como o valor de um imóvel ou a temperatura.