O que faz a instrução

X\_train, X\_test, y\_train, y\_test = train\_test\_split( X, y, test\_size=0.3, random\_state=42, stratify=y )

Essa instrução:

X\_train, X\_test, y\_train, y\_test = train\_test\_split(

X, y, test\_size=0.3, random\_state=42, stratify=y

)

faz **a divisão dos seus dados em conjuntos de treino e teste**, que é uma etapa essencial em Machine Learning. Vamos detalhar cada parte:

**1️ train\_test\_split**

* É uma função do scikit-learn que **divide arrays ou matrizes em conjuntos de treino e teste**.

**2️ Parâmetros**

* X → suas **features** (as informações de entrada do modelo).
* y → seus **rótulos/targets** (o que você quer prever).
* test\_size=0.3 → **30% dos dados vão para teste**, 70% para treino.
* random\_state=42 → garante que a divisão seja **reproduzível**; sempre que você rodar com o mesmo número, a divisão será igual.
* stratify=y → garante que a proporção das classes em y seja **mantida tanto no treino quanto no teste**, evitando que uma classe fique subrepresentada.

**3️ O que a linha retorna**

* X\_train → features usadas para **treinar o modelo**.
* X\_test → features usadas para **testar o modelo** (verificar se aprendeu corretamente).
* y\_train → rótulos correspondentes ao X\_train.
* y\_test → rótulos correspondentes ao X\_test.

💡 **Resumo visual mental:**

Todos os dados (X, y)

│

├── 70% treino → X\_train, y\_train

└── 30% teste → X\_test, y\_test