Avaliação de modelos:

1. **Validação simples**: separar o dataset em treinamento e validação. Criar o modelo a partir do conjunto de treinamento e validar o modelo no conjunto de validação.
2. **Cross-validation (k-fold cross validation)**: considerando todo o dataset. Criar modelos e validar usando k-fold cross validation.
3. **Cross-validation para o dataset de treinamento com dataset de validação**: separar o dataset em treinamento e validação. Criar diversos modelos e validá-los usando cross-validation sobre o conjunto de treinamento. Ao final, quando tivermos apenas um modelo (o melhor modelo levando-se em consideração o conjunto de treinamento) então podemos validar o modelo usando o dataset de validação.

Problemas com a utilização de modelos preditivos:

1. **Overfitting**: quando o erro do modelo no conjunto de treinamento é baixo, porém o erro do modelo no conjunto de teste é alto. Como resolver problemas de overfitting?
   1. Adicionar mais exemplos ao conjunto de treinamento
   2. Tornar o modelo menos complexo: removendo atributos ou escolhendo um algoritmo mais simples.
2. **Underfitting**: quando o erro do modelo é alto em ambos os casos. Como resolver problemas de underfitting?
   1. Tornar o modelo mais complexo. E isto pode acontecer de duas maneiras: escolhendo um algoritmo mais complexo ou adicionando mais atributos ao dataset.

Quando o modelo funciona? Quando o erro é baixo para ambos os casos.

O que é erro baixo? Depende do objetivo do projeto. A definição de erro aceitável deve ser feita no início do projeto como parte da definição do escopo.