

Classificação

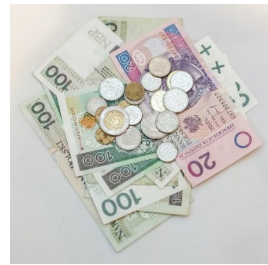


Queremos prever ou descrever a classe de um evento



Normalmente a classe em uma relação esta representada em um atributo especial, posicionado como último atributo da relação

Exemplos de Classificação



Modelo

Palavra	Spam	Não Spam
oferta	0,9	0,1
ganhe	0,8	0,2
preço	0,3	0,7
grátis	0,7	0,3



Medir o Desempenho do Modelo



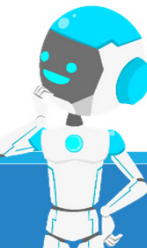
Treino: Algoritmo processa dados e cria modelo



Validação: Dados são usados para ajustar o modelo

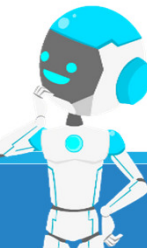
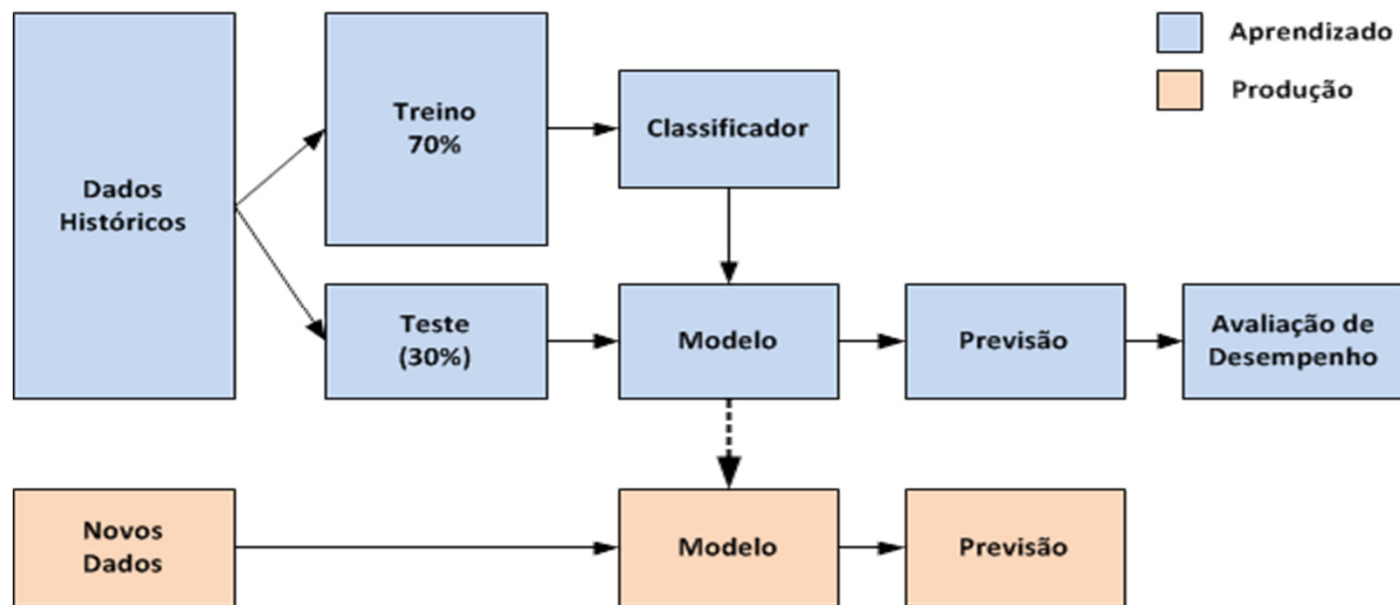


Teste: Dados são usados para avaliar a performance do modelo



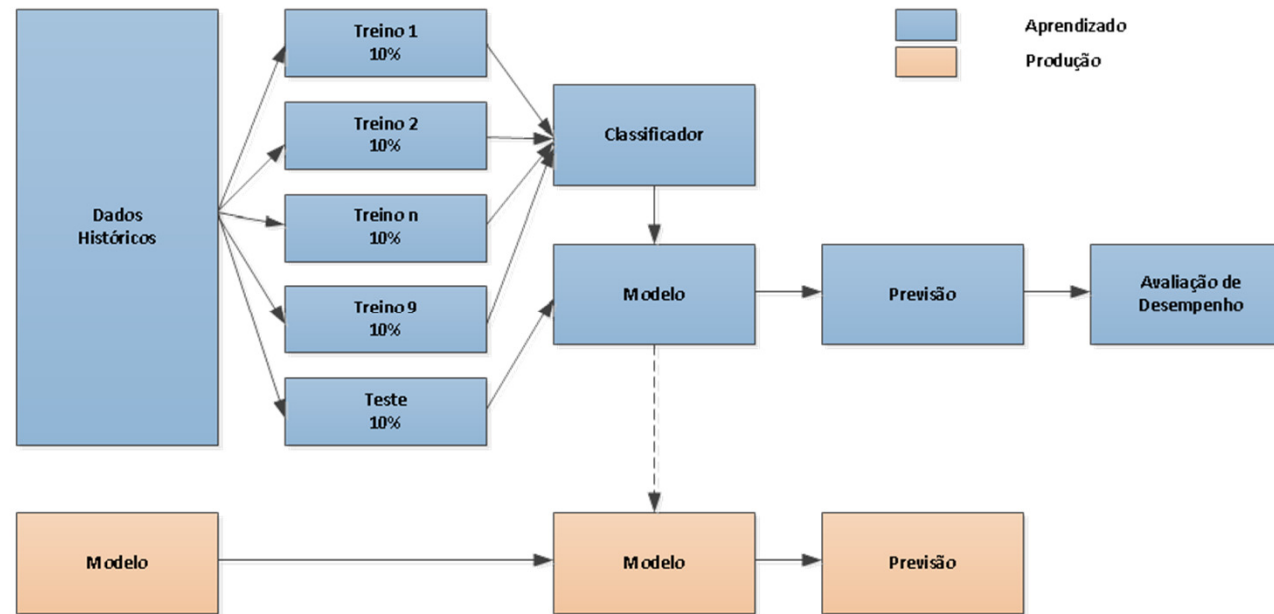
Hold-Out

Divide os dados em treino e teste



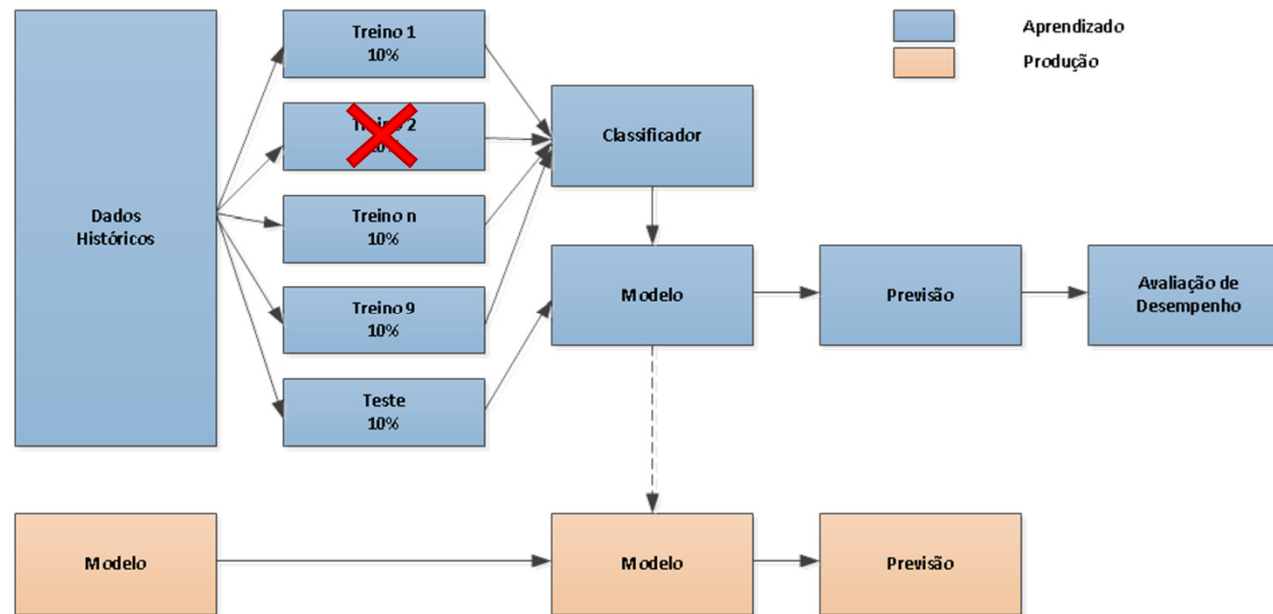
Validação Cruzada

Divide os dados em vários conjuntos menores



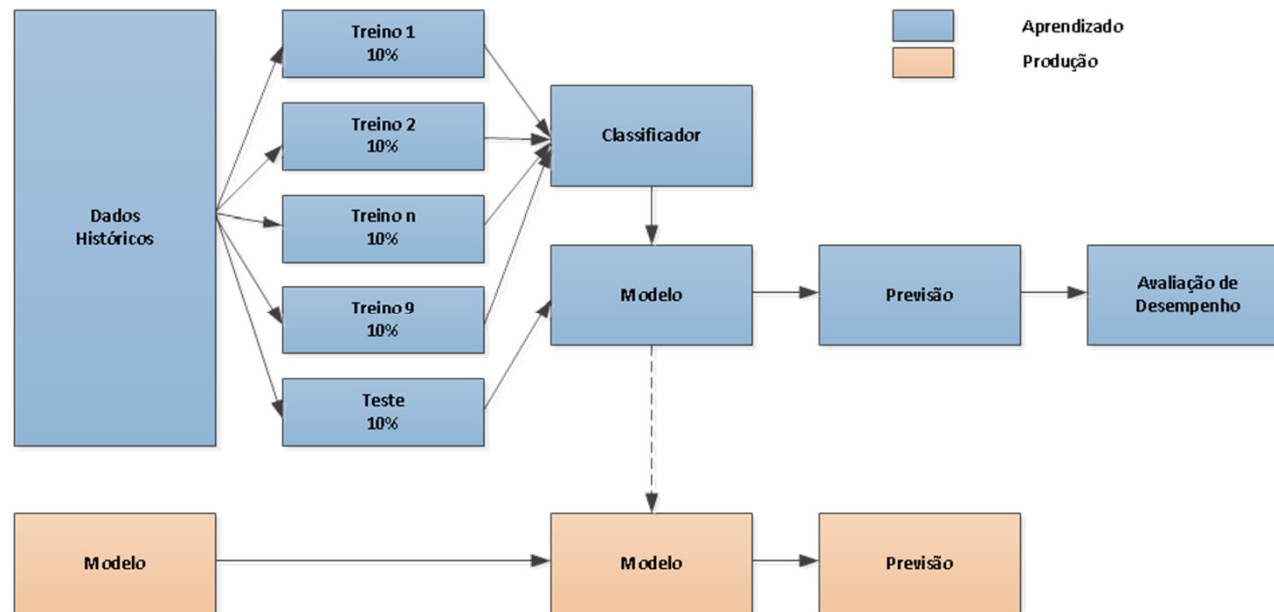
Leave-One-Out

Tipo de validação cruzada, que treina com todos os conjuntos de dados menos um



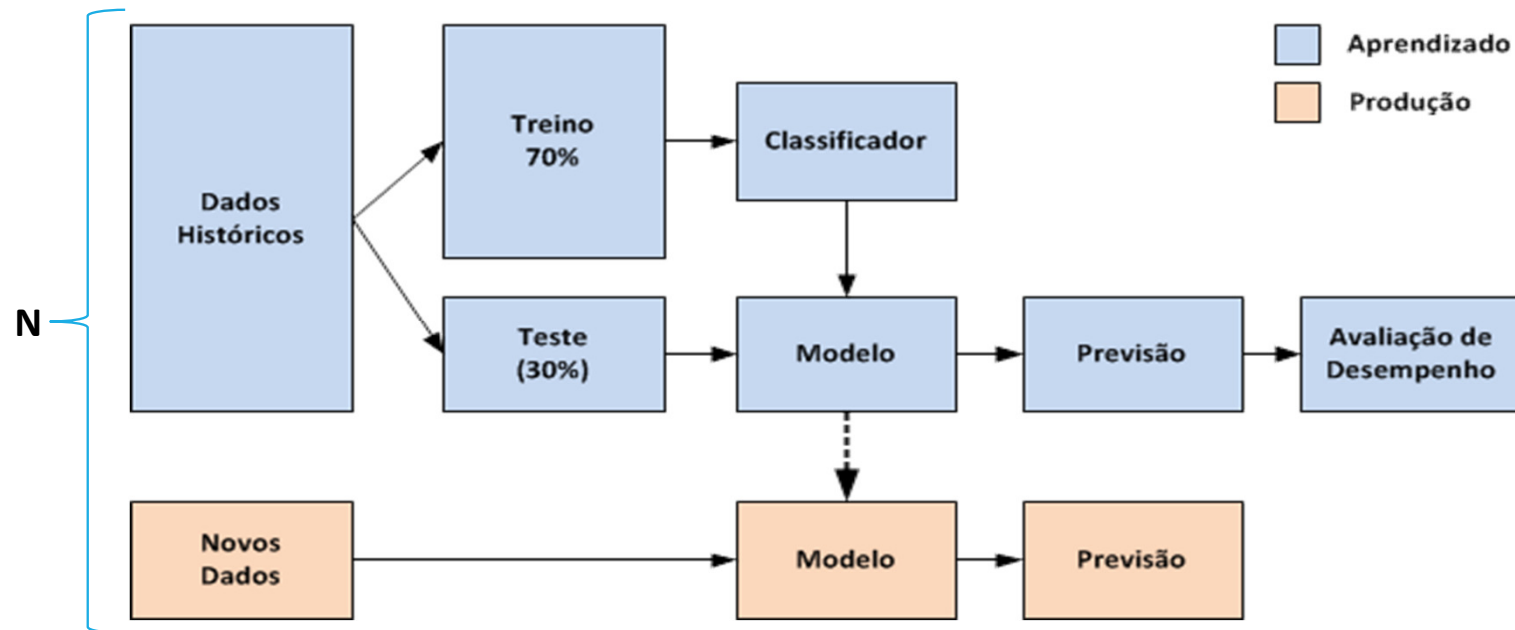
K-Fold

Tipo de validação Cruzada: Divide o conjunto de dados em k subconjuntos, treinar o modelo em k-1 subconjuntos e avaliá-lo no subconjunto restante. Esse processo é repetido k vezes,



Subamostragem

É uma técnica que envolve a seleção de uma amostra aleatória de exemplos do conjunto de dados original para treinamento do modelo



Generalização Versus Super Ajuste Versus Sub Ajuste

❖ O objetivo de todo classificador é criar modelos genéricos



Genérico
Ajustado



Super/Sub

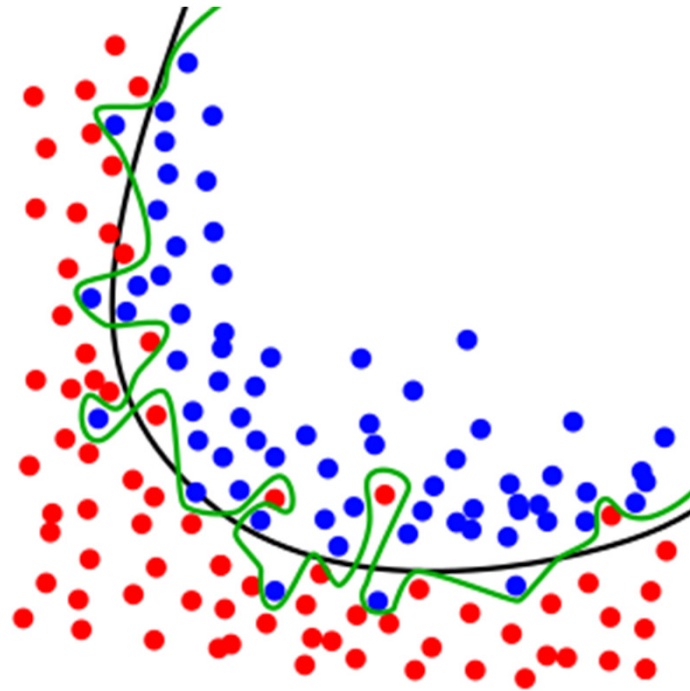


Super Ajuste

O modelo super ajustado funciona bem com dados de treino, mas tem o desempenho pobre em dados de teste ou de produção.



Super Ajuste



Causas de Super Ajuste



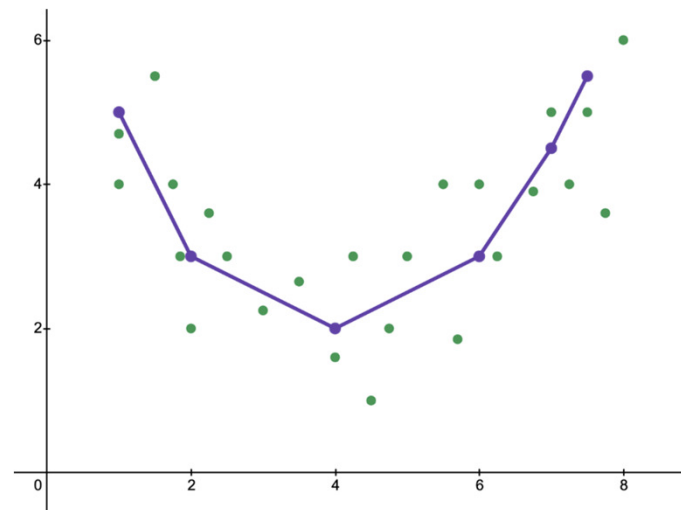
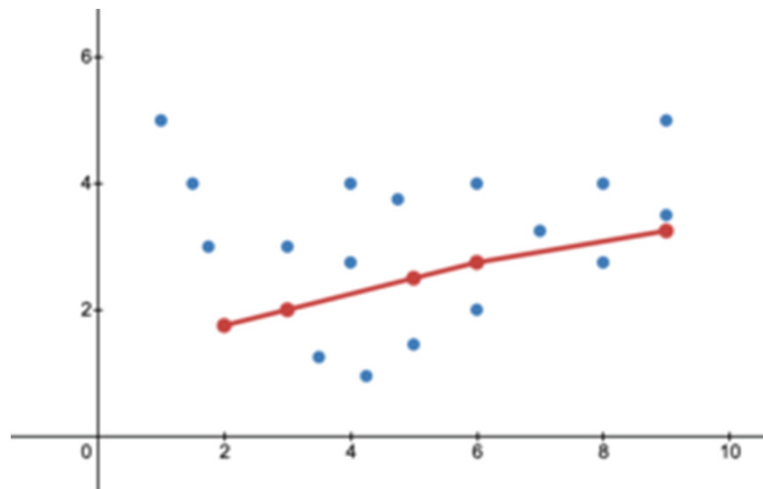
- Tamanho insuficiente do conjunto de dados
- Complexidade excessiva do modelo de treinamento
- Ruído nos dados de treinamento
- Seleção inadequada de atributos
- Falta de validação cruzada

Sub Ajuste

Modelo de machine learning não consegue se ajustar bem aos dados de treinamento e, portanto, também não consegue generalizar bem para novos dados.



Sub Ajuste



Causas do Sub Ajuste

1. Modelo muito simples
2. Conjunto de dados muito pequeno
3. Seleção inadequada de atributos
4. Falta de ajuste de hiperparâmetros

