

# Classificação com k-NN

Minicurso Ciência de Dados

# Objetivos da Aula

- Entender o conceito de **classificação**
- Conhecer o algoritmo **k-Vizinhos Mais Próximos (k-NN)**
- Aplicar o k-NN em um conjunto de dados fictício
- Avaliar o desempenho do modelo

# O que é Classificação?

- Uma tarefa de aprendizado supervisionado
- Objetivo: **prever a classe** de uma nova amostra com base em exemplos anteriores
- Exemplos:
  - Detectar se um e-mail é spam ou não
  - Classificar frutas (maçã, banana, laranja)
  - Diagnosticar doenças a partir de sintomas

# Algoritmo k-Vizinhos Mais Próximos (k-NN)

- Baseia-se na **proximidade** entre amostras
- Passos:
  - 1. Escolher o número  $k$  (quantos vizinhos considerar)
  - 2. Calcular a distância entre a amostra nova e as amostras do treino
  - 3. Ver qual classe é mais frequente entre os  $k$  vizinhos mais próximos
- Vantagens: Simples, interpretável
- Limitações: Pode ser lento em grandes bases

# Fluxo da Tarefa de Classificação

- 1. Coletar dados
- 2. Dividir em treino e teste
- 3. Normalizar (opcional, mas recomendável)
- 4. Treinar o modelo
- 5. Avaliar com métricas (acurácia, matriz de confusão)
- 6. Fazer previsões

# Métricas de Avaliação

- **Acurácia:** porcentagem de acertos
- **Matriz de confusão:** mostra erros por classe
- **Precision, Recall, F1-score:** úteis quando classes estão desbalanceadas

# Atividade Prática

- Abrir o notebook fornecido
- Executar cada célula
- Alterar o valor de  $k$  e observar mudanças
- Testar novas frutas com valores fictícios
- Verificar o impacto na acurácia