



Classificador

- Redes Neurais
- Máquina de Vetor de Suporte
- Naive Bayes
- Árvores de Decisão
- Regressão Linear
- **Florestas Aleatórias**



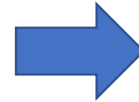
Transformação de Dados

- Computadores são capazes de processar apenas números!
- Dados categóricos precisam ser transformados!
- *Categorical encoding é o processo de transformar categorias em números*
- *Duas Formas:*
 - Label encoding
 - One-hot encoding

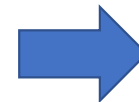
Label encoding

- Cada categoria recebe um número, normalmente em ordem alfabética

EstadoCivil
Casado
Solteiro
Divorciado
Casado
Solteiro
Casado
Solteiro
Casado
Casado
Solteiro
Casado
Solteiro
Divorciado
Casado
Solteiro
Casado
Casado
Casado
Casado
Solteiro



Casado	0
Divorciado	1
Solteiro	2

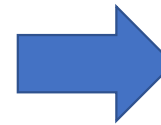


EstadoCivil
0
1
2
0
2
0
2
0
0
2
0
2
1
0
2
0
0
0
0
2

One-hot encoding

- Cada categoria é transformada em outro atributo: dummy variable
- Um valor binário informa a ocorrência

EstadoCivil
Casado
Solteiro
Divorciado
Casado
Solteiro
Casado
Solteiro
Casado
Casado
Solteiro
Casado
Solteiro
Divorciado
Casado
Solteiro
Casado
Casado
Casado
Solteiro



Casado	Solteiro	Divorciado
1	0	0
0	1	0
0	0	1
1	0	0
0	1	0
1	0	0
0	1	0
1	0	0
1	0	0
0	1	0
1	0	0
0	1	0
0	0	1
1	0	0
0	1	0
1	0	0
1	0	0
1	0	0
0	1	0

Dummy Variable Trap

- O valor dos atributos se torna altamente previsível
- Resultado, correlação entre as variáveis Independentes: multicolinearidade
- Solução: Excluir um dos atributos!

Casado	Solteiro	Divorciado
1	0	0
0	1	0
0	0	1
1	0	0
0	1	0
1	0	0
0	1	0
1	0	0
1	0	0
0	1	0
1	0	0
0	1	0
0	0	1
1	0	0
0	1	0
1	0	0
1	0	0
1	0	0
1	0	0
0	1	0

Scikit-Learn

- Ampla biblioteca de Machine Learning
- Popular, eficiente e madura
- Open source
- Pode ser utilizada Comercialmente