



# MACHINE LEARNING APLICADO À ANÁLISE DE DADOS

AULA 7 – 24/03/2021

ÁRVORE DE DECISÃO

# ÁRVORE DE DECISÃO (DECISION TREE)

A árvore de decisão é um tipo de algoritmo de aprendizagem de máquina supervisionada. São utilizados em tarefas de Classificação e Regressão. O objetivo do algoritmo é, basicamente, criar um modelo que preveja o valor de uma variável target (alvo), aprendendo regras de decisão simples inferidas dos atributos dos dados.

# O QUE DEVEMOS SABER ANTES DE USAR O ALGORITMO?

- Fácil entendimento e interpretação
- Aceita dados categóricos e numéricos
- Requer pouca preparação de dados
- Não oferece suporte a valores ausentes (dados missing)

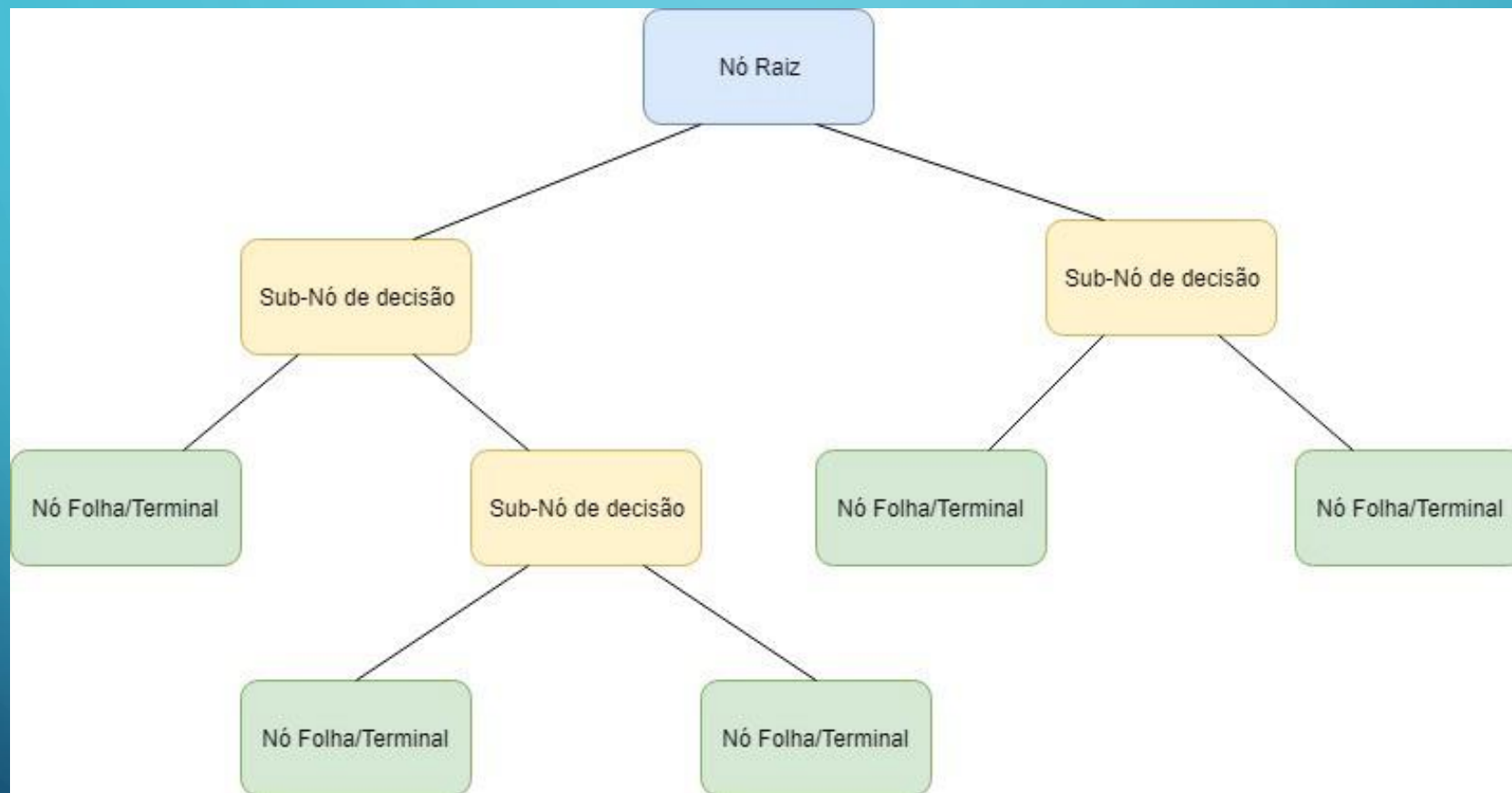
# O QUE DEVEMOS SABER ANTES DE USAR O ALGORITMO?

- É possível validar um modelo por meio de testes estatísticos
- É propensa a sofrer Overfitting (sobreajuste)
- São instáveis: pequenas alterações no dados de treino podem modificar a árvore

# ALGUMAS ESTRATÉGIAS PARA AJUDAR A EVITAR O OVERFITTING:

- Especificar o número mínimo de amostras para divisão do nó. Valores altos previnem a criação de modelos complexos.
- Especificar o número mínimo de amostras para o nível folha. Nos permite controlar o crescimento da árvore.
- Definir a profundidade máxima da árvore.
- Número máximo de features(atributos, colunas) para considerar durante a divisão.

# ESTRUTURA DE UMA ÁRVORE DE DECISÃO



(Árvore de Decisão em R. Classificação e Aplicação na Base de... | by Gabriel Stankevix | Medium)

# TERMINOLOGIAS

- Nó raiz: é o atributo do dataset que melhor divide o conjunto de dados
- Nós de decisão: são representadas pelas condições lógicas (IF/ELSE) que determinam o caminho dentro da árvore
- Nós folhas: onde é atribuído a **classe**, no caso de **Classificação** e a **média** das observações, no caso da **Regressão**

The background is a blue gradient with decorative white circuit-like lines in the corners. These lines consist of straight segments and small circles, resembling a stylized electronic circuit board.

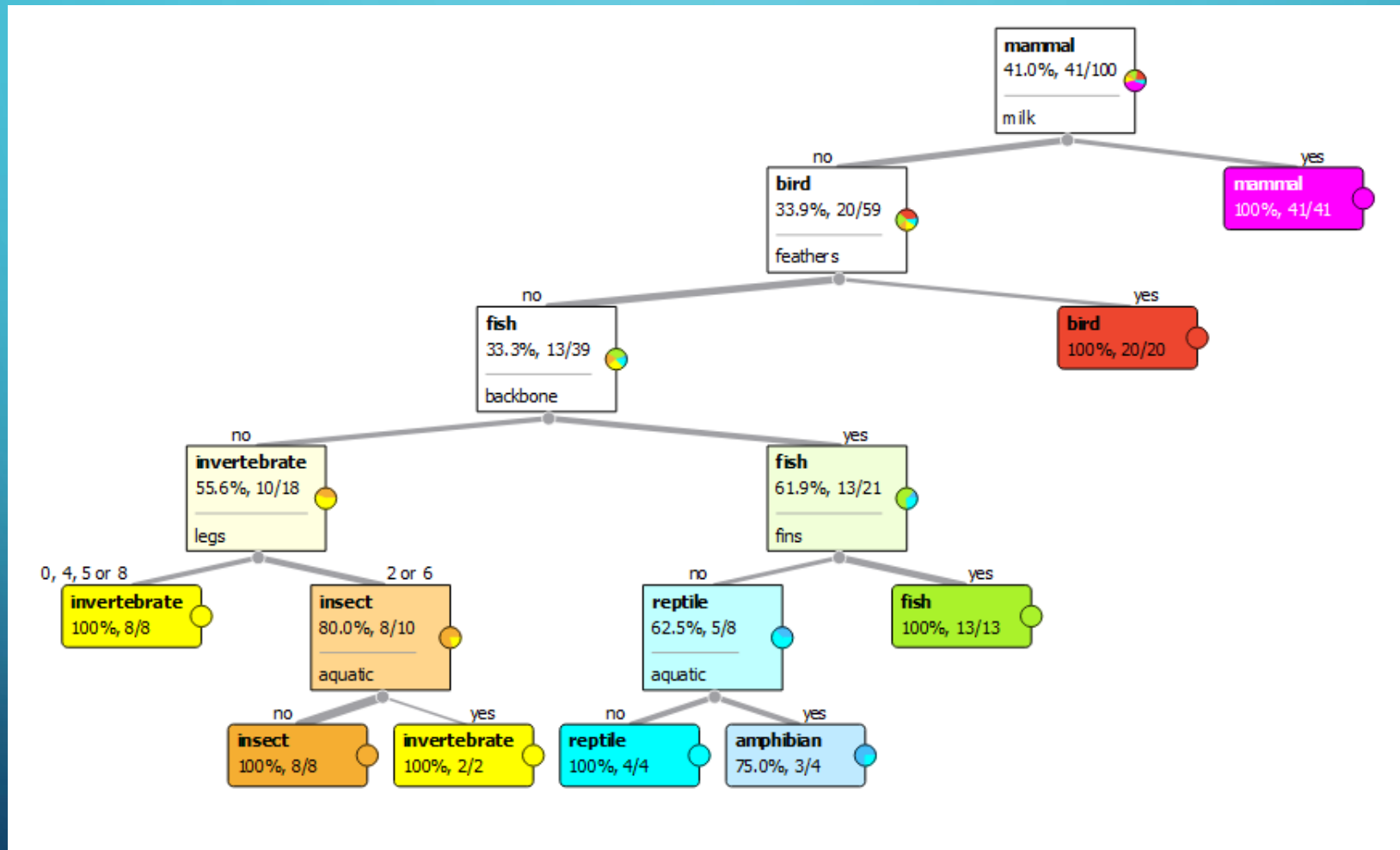
EXEMPLO NO ORANGE



## PARTE DO DATASET ZOO (UCI)

	type	name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone
1	mammal	aardvark	yes	no	no	yes	no	no	yes	yes	yes
2	mammal	antelope	yes	no	no	yes	no	no	no	yes	yes
3	fish	bass	no	no	yes	no	no	yes	yes	yes	yes
4	mammal	bear	yes	no	no	yes	no	no	yes	yes	yes
5	mammal	boar	yes	no	no	yes	no	no	yes	yes	yes
6	mammal	buffalo	yes	no	no	yes	no	no	no	yes	yes
7	mammal	calf	yes	no	no	yes	no	no	no	yes	yes
8	fish	carp	no	no	yes	no	no	yes	no	yes	yes
9	fish	catfish	no	no	yes	no	no	yes	yes	yes	yes
10	mammal	cavy	yes	no	no	yes	no	no	no	yes	yes
11	mammal	cheetah	yes	no	no	yes	no	no	yes	yes	yes
12	bird	chicken	no	yes	yes	no	yes	no	no	no	yes
13	fish	chub	no	no	yes	no	no	yes	yes	yes	yes
14	invertebrate	clam	no	no	yes	no	no	no	yes	no	no
15	invertebrate	crab	no	no	yes	no	no	yes	yes	no	no
16	invertebrate	crayfish	no	no	yes	no	no	yes	yes	no	no
17	bird	crow	no	yes	yes	no	yes	no	yes	no	yes
18	mammal	deer	yes	no	no	yes	no	no	no	yes	yes
19	fish	dogfish	no	no	yes	no	no	yes	yes	yes	yes
20	mammal	dolphin	no	no	no	yes	no	yes	yes	yes	yes
21	bird	dove	no	yes	yes	no	yes	no	no	no	yes
22	bird	duck	no	yes	yes	no	yes	yes	no	no	yes
23	mammal	elephant	yes	no	no	yes	no	no	no	yes	yes
24	bird	flamingo	no	yes	yes	no	yes	no	no	no	yes
25	insect	flea	no	no	yes	no	no	no	no	no	no
26	amphibian	frog	no	no	yes	no	no	yes	yes	yes	yes
27	mammal	fruitbat	yes	no	no	yes	yes	no	no	yes	yes
28	mammal	giraffe	yes	no	no	yes	no	no	no	yes	yes
29	mammal	girl	yes	no	no	yes	no	no	yes	yes	yes
30	insect	gnat	no	no	yes	no	yes	no	no	no	no
31	mammal	goat	yes	no	no	yes	no	no	no	yes	yes
32	mammal	gorilla	yes	no	no	yes	no	no	no	yes	yes
33	bird	gull	no	yes	yes	no	yes	yes	yes	no	yes
34	fish	haddock	no	no	yes	no	no	yes	no	yes	yes
35	mammal	hamster	yes	no	no	yes	no	no	no	yes	yes

## VISUALIZAÇÃO DO TREE VIEWER COM O DATASET ZOO (ULC) NO SOFTWARE ORANGE



# REFERÊNCIAS BIBLIOGRÁFICAS + LINKS ÚTEIS

- [Arvore de Decisão em R. Classificação e Aplicação na Base de... | by Gabriel Stankevix | Medium](https://medium.com/@gabriel.stankevix/arvore-de-decis%C3%A3o-em-r-85a449b296b2) (<https://medium.com/@gabriel.stankevix/arvore-de-decis%C3%A3o-em-r-85a449b296b2>)
- [1.10. Decision Trees — scikit-learn 0.24.1 documentation](https://scikit-learn.org/stable/modules/tree.html) (<https://scikit-learn.org/stable/modules/tree.html>)
- [Minerando Dados › Árvores de Decisão](https://minerandodados.com.br/arvores-de-decisao-conceitos-e-aplicacoes/) (<https://minerandodados.com.br/arvores-de-decisao-conceitos-e-aplicacoes/>)
- [Orange Data Mining - Data Mining](https://orangedatamining.com/) (<https://orangedatamining.com/>)