

Árvore de Decisão

Agente Educacional

Sérgio M. Dias

Agenda

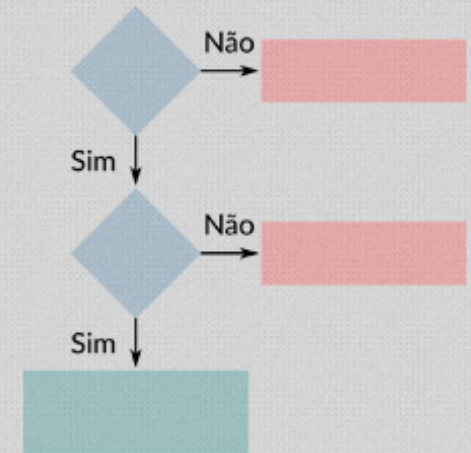
- Árvore de decisão
- Treinamento
- Alguns aspectos de ADs
- Algoritmos para ADs
- Vantagens e Desvantagens de ADs
- Exercício KNIME AD



Árvore de Decisão - AD

Forma mais simples

- Lista de perguntas (respostas “*sim*” ou “*não*”)
- Hierarquicamente *arranjadas*
- Levam a uma *decisão*



Estrutura da árvore determinada por meio de aprendizado

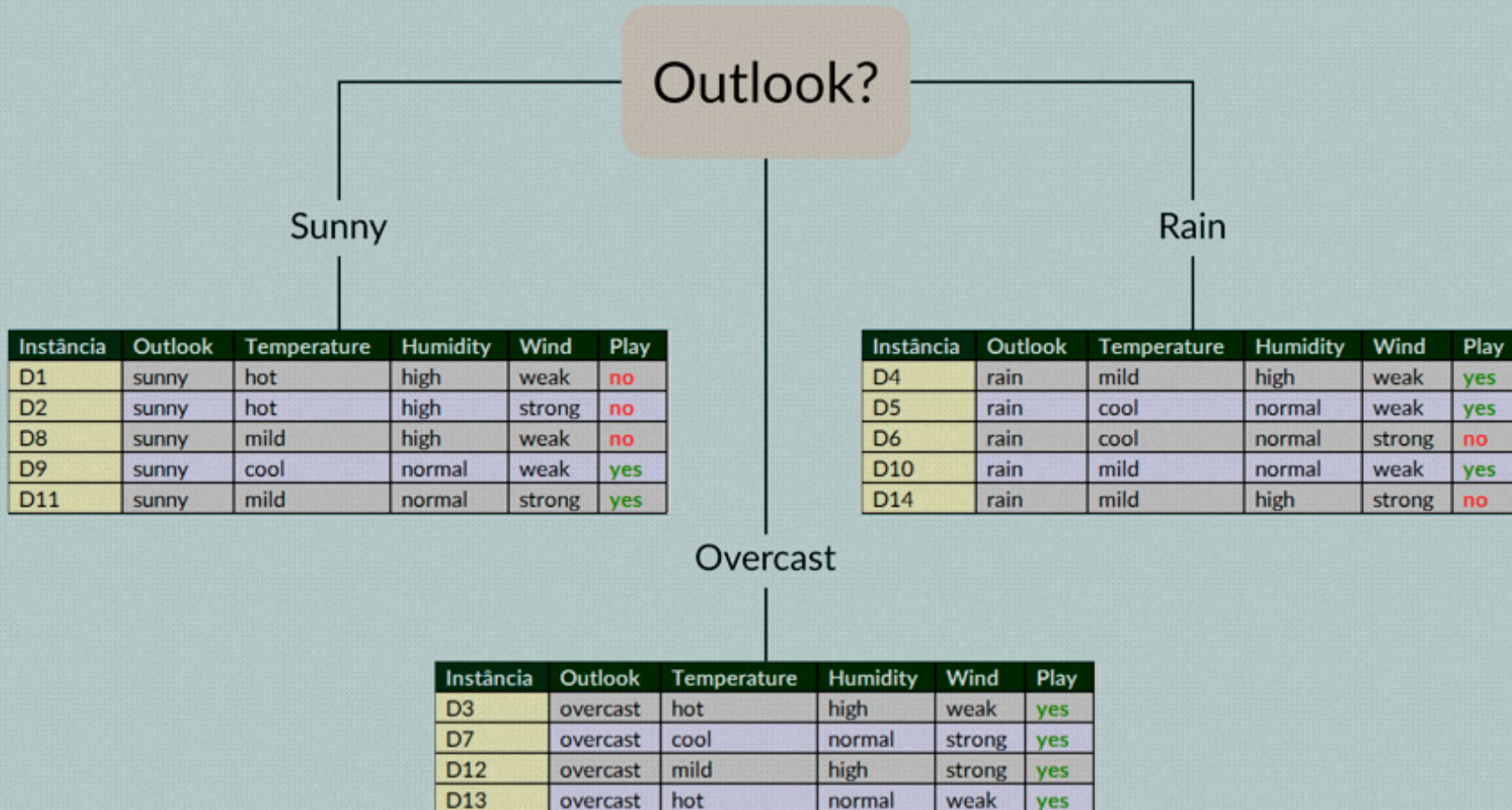
Treinamento

AD encontra regras que **recursivamente bifurcam**, baseadas nos valores dos atributos, o conjunto de dados.

Conteúdo dos subconjuntos pode ser **descrito por um conjunto de regras**.

Base de Dados “Tempo”

Instância	Outlook	Temperature	Humidity	Wind	Play
D1	sunny	hot	high	weak	no
D2	sunny	hot	high	strong	no
D3	overcast	hot	high	weak	yes
D4	rain	mild	high	weak	yes
D5	rain	cool	normal	weak	yes
D6	rain	cool	normal	strong	no
D7	overcast	cool	normal	strong	yes
D8	sunny	mild	high	weak	no
D9	sunny	cool	normal	weak	yes
D10	rain	mild	normal	weak	yes
D11	sunny	mild	normal	strong	yes
D12	overcast	mild	high	strong	yes
D13	overcast	hot	normal	weak	yes
D14	rain	mild	high	strong	no



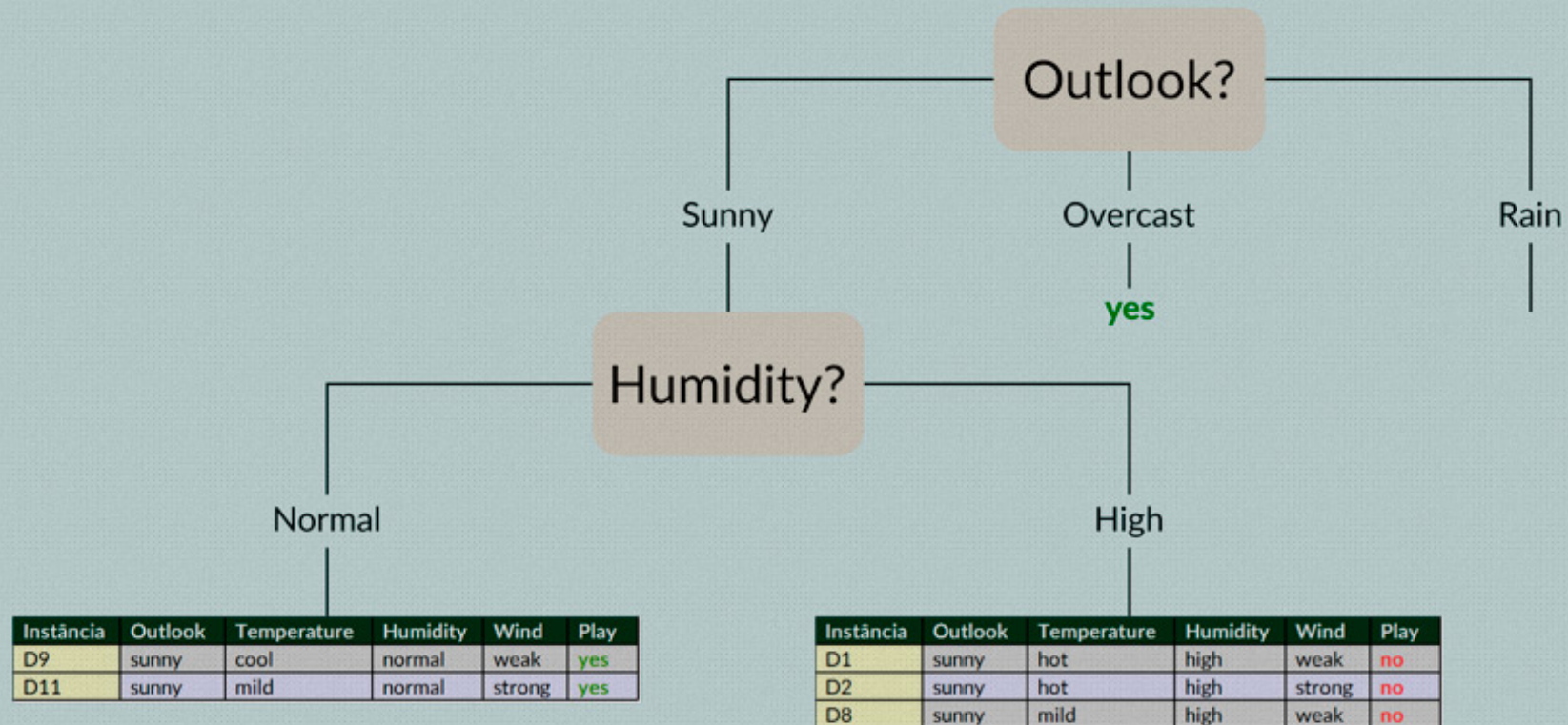
Teste

if outlook = sunny

if outlook = overcast

if outlook = rain

Exemplo	Outlook	Temperature	Humidity	Wind	Play?
D1	sunny	hot	high	weak	no
D2	sunny	hot	high	strong	no
D8	sunny	mild	high	weak	no
D9	sunny	cool	normal	weak	yes
D11	sunny	mild	normal	strong	yes
D3	overcast	hot	high	weak	yes
D7	overcast	cool	normal	strong	yes
D12	overcast	mild	high	strong	yes
D13	overcast	hot	normal	weak	yes
D4	rain	mild	high	weak	yes
D5	rain	cool	normal	weak	yes
D6	rain	cool	normal	strong	no
D10	rain	mild	normal	weak	yes
D14	rain	mild	high	strong	no



Teste

if outlook = sunny
and humidity = high



Exemplo	Outlook	Temperature	Humidity	Wind	Play?
D1	sunny	hot	high	weak	no
D2	sunny	hot	high	strong	no
D8	sunny	mild	high	weak	no

if outlook = sunny
and humidity = normal



D9	sunny	cool	normal	weak	yes
D11	sunny	mild	normal	strong	yes

if outlook = overcast



D3	overcast	hot	high	weak	yes
D7	overcast	cool	normal	strong	yes
D12	overcast	mild	high	strong	yes
D13	overcast	hot	normal	weak	yes

if outlook = rain
and wind = strong

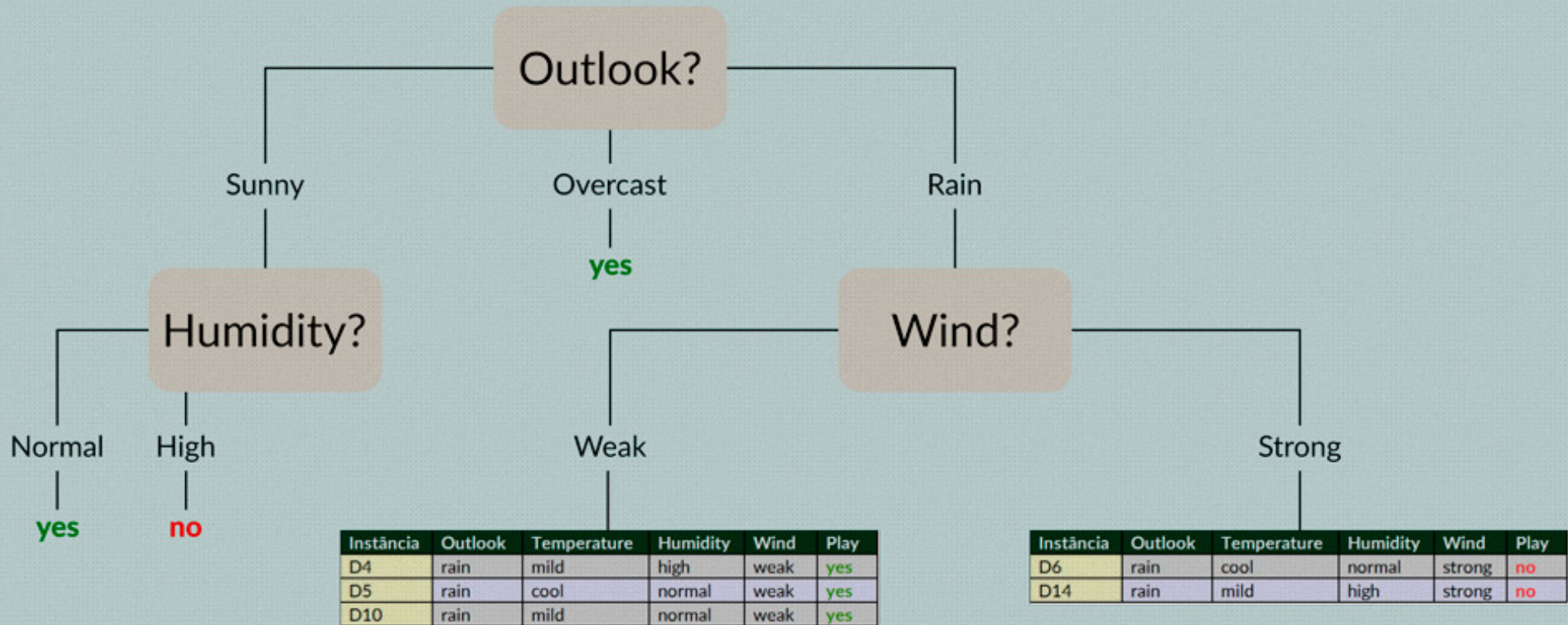


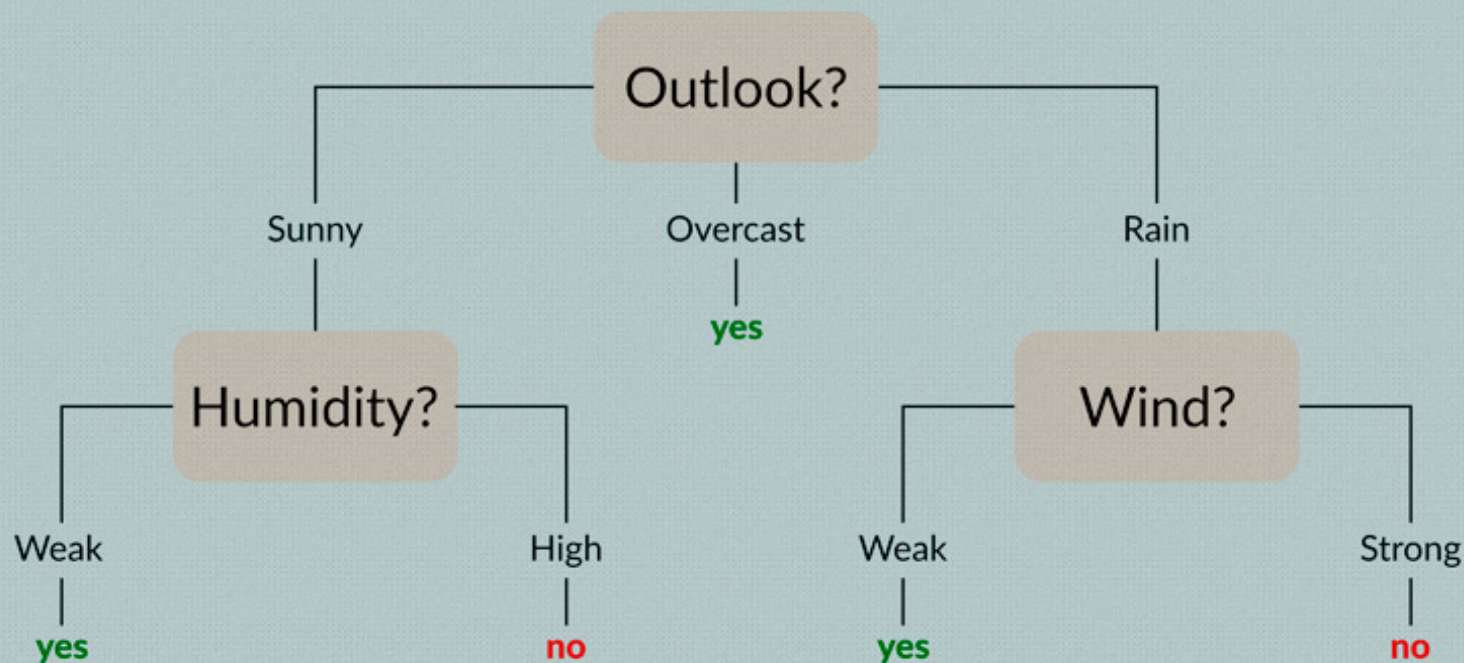
D6	rain	cool	normal	strong	no
D14	rain	mild	high	strong	no

if outlook = rain
and wind = weak



D4	rain	mild	high	weak	yes
D5	rain	cool	normal	weak	yes
D10	rain	mild	normal	weak	yes





Teste

if outlook = sunny
and humidity = high



Exemplo	Outlook	Temperature	Humidity	Wind	Play?
D1	sunny	hot	high	weak	no
D2	sunny	hot	high	strong	no
D8	sunny	mild	high	weak	no

if outlook = sunny
and humidity = normal



D9	sunny	cool	normal	weak	yes
D11	sunny	mild	normal	strong	yes

if outlook = overcast



D3	overcast	hot	high	weak	yes
D7	overcast	cool	normal	strong	yes
D12	overcast	mild	high	strong	yes
D13	overcast	hot	normal	weak	yes

if outlook = rain
and wind = strong



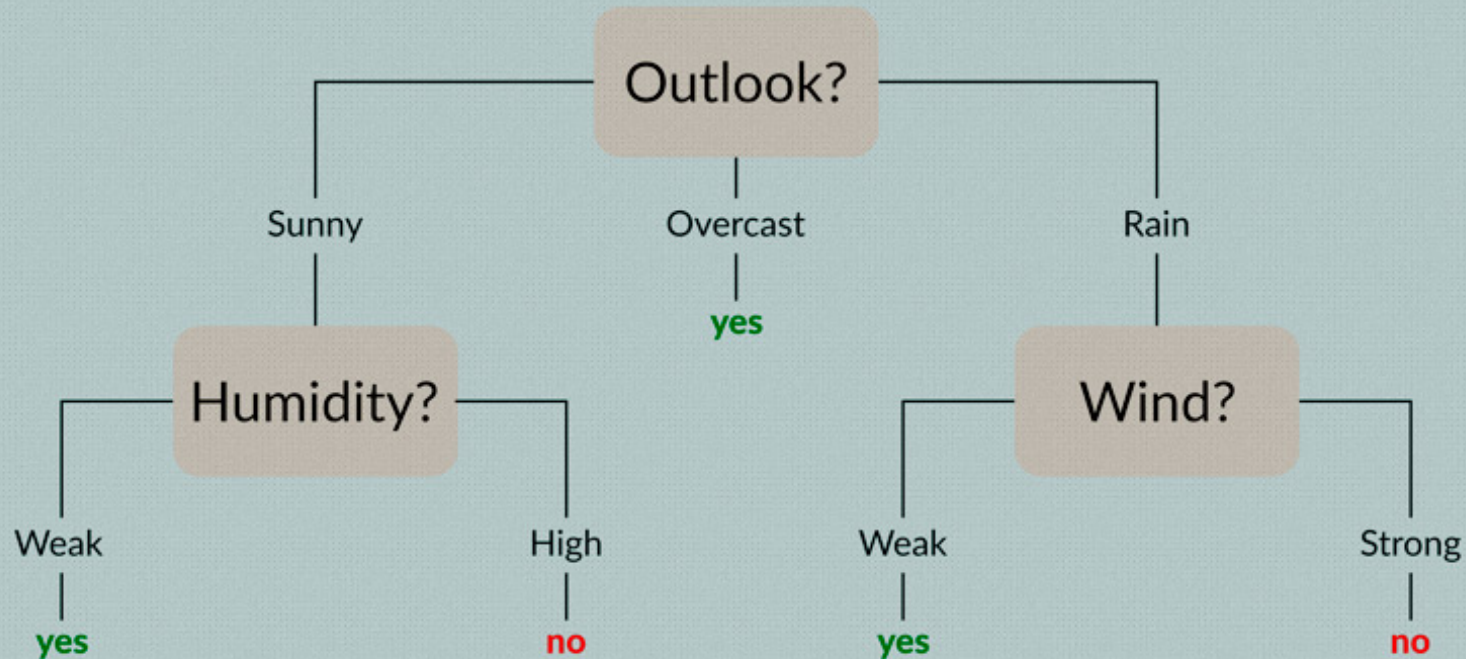
D6	rain	cool	normal	strong	no
D14	rain	mild	high	strong	no

if outlook = rain
and wind = weak



D4	rain	mild	high	weak	yes
D5	rain	cool	normal	weak	yes
D10	rain	mild	normal	weak	yes

Como classificar <rain, hot, high, normal, strong>?



Alguns Aspectos de ADs

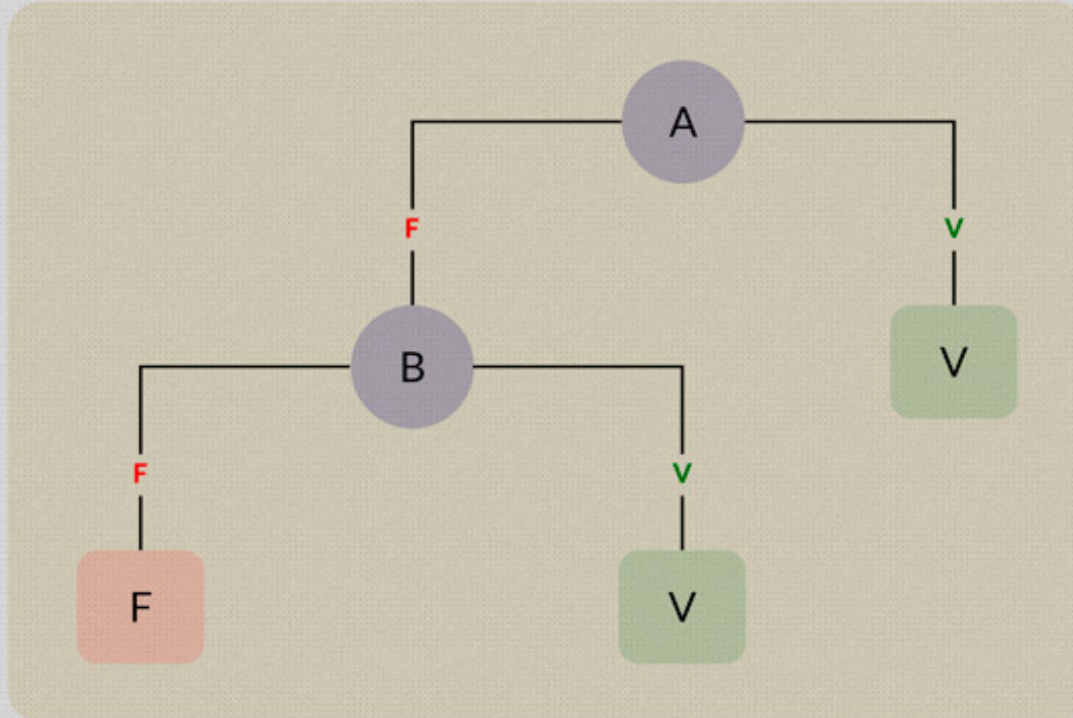
Existem critérios para **escolha dos atributos** em cada posição da árvore
Entropia, ganho de informação, etc.

A árvore pode ficar **muito grande** (muitos passos entre a raiz e uma folha)
Os algoritmos adotam critérios de **poda para reduzir esse percurso.**

Cada ramificação é uma **regra.**

A árvore representa conhecimento com **junções** e **disjunções** (E e OU).

Alguns Aspectos de AD



A	B	A+B
F	F	F
F	V	V
V	F	V
V	V	V

Algoritmos para ADs

O ID3

- Primeiro algoritmo em **indução** de árvores de decisão
- **Recursivo** e baseado em **busca gulosa**, procurando, sobre um conjunto de atributos, aqueles que “melhor” dividem os exemplos, gerando subárvores

C4.5

- Evolução do ID3
- Lida tanto com atributos **categóricos** (ordinais ou não ordinais) como com atributos **contínuos**

CART (*Classification and Regression Trees*)

- As árvores geradas pelo algoritmo **CART** são sempre **binárias**, as quais podem ser percorridas da sua raiz até as folhas, respondendo apenas a **questões simples** do tipo “**sim**” ou “**não**”

Vantagens e Desvantagens de ADs

Vantagens

- Estrutura de fácil manipulação
- Produzem modelos que podem ser facilmente interpretados e visualizados por humanos

Desvantagens

- Pouca robustez para lidar com dados de grande dimensão
- Acurácia afetada por atributos pouco relevantes
- Dificuldade em lidar com dados contínuos

<Exercício **KNIME AD**>



Obrigado!

Agente Educacional

Sérgio M. Dias

sergio.dias@serpro.gov.br | #31 6539

Demais agentes educacionais sobre o assunto

Marcelo Pita | marcelo.pita@serpro.gov.br | #81 8794

Gustavo Torres | gustavo.gamatorres@serpro.gov.br | #31 6950