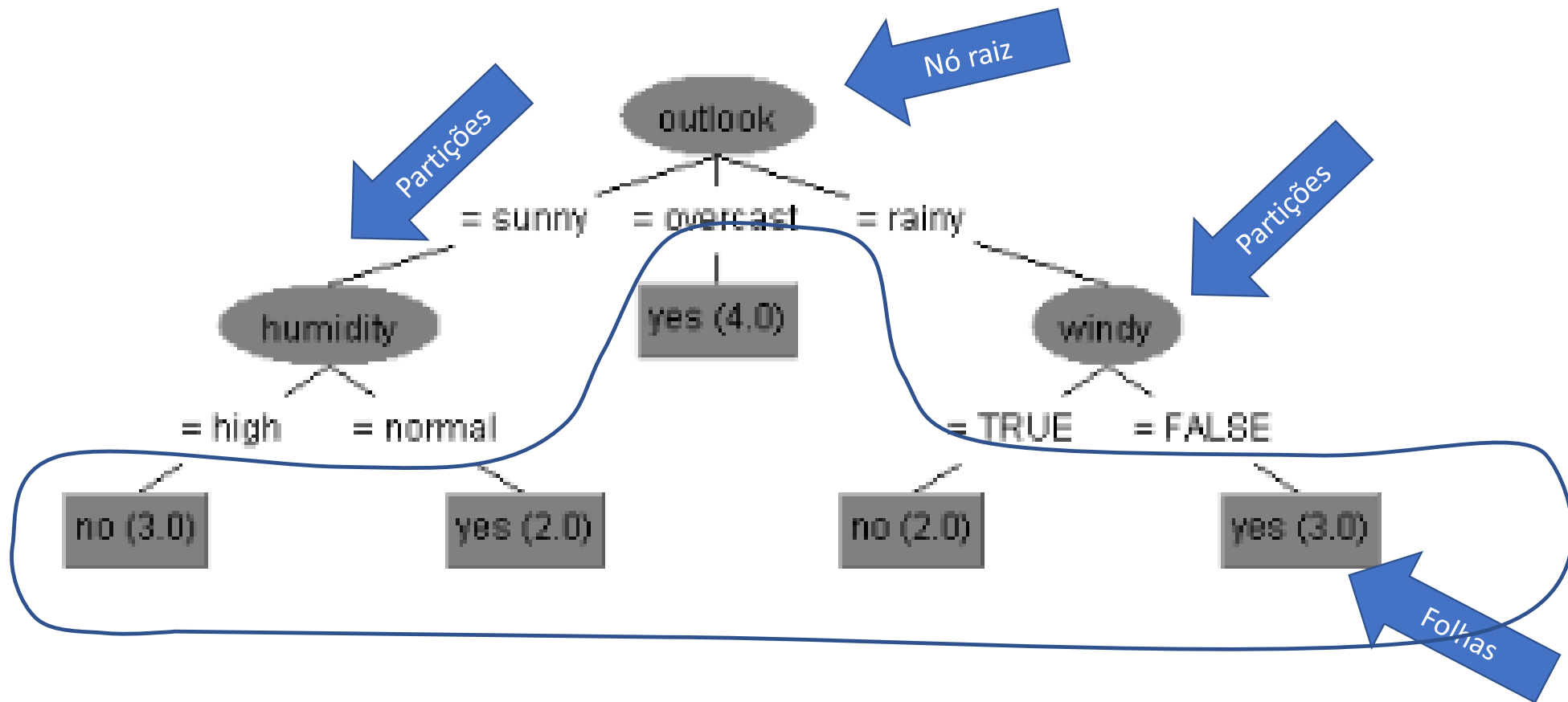
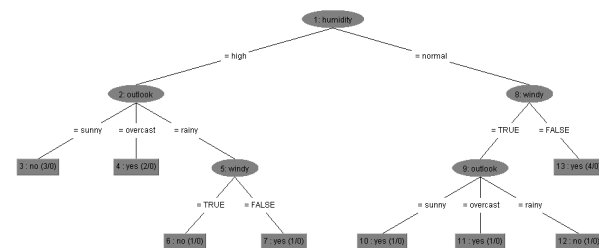
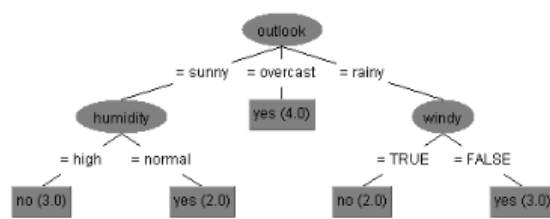
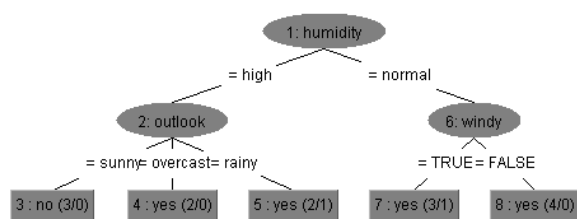


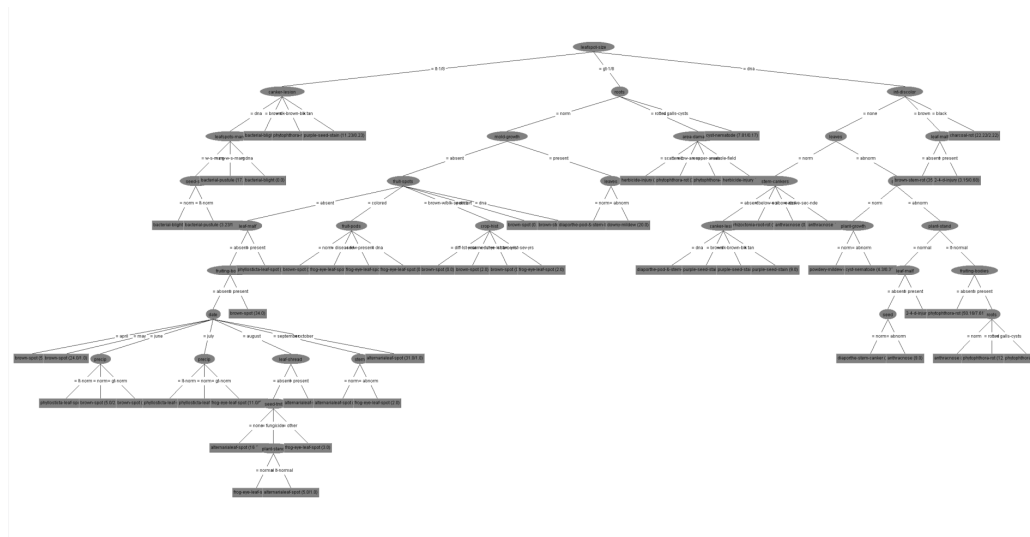
Estrutura





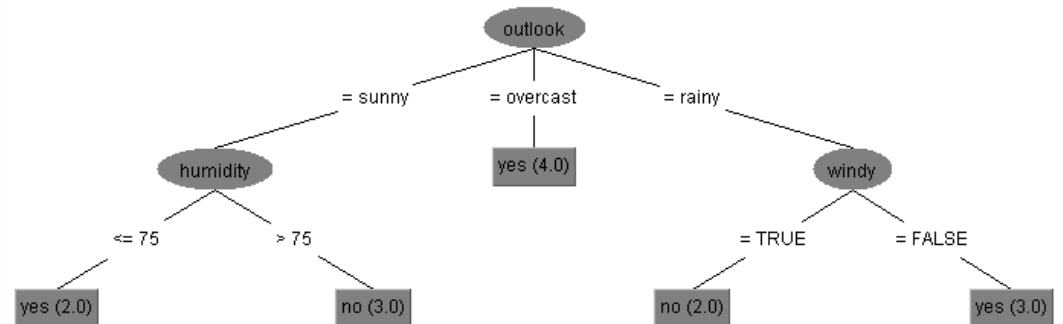
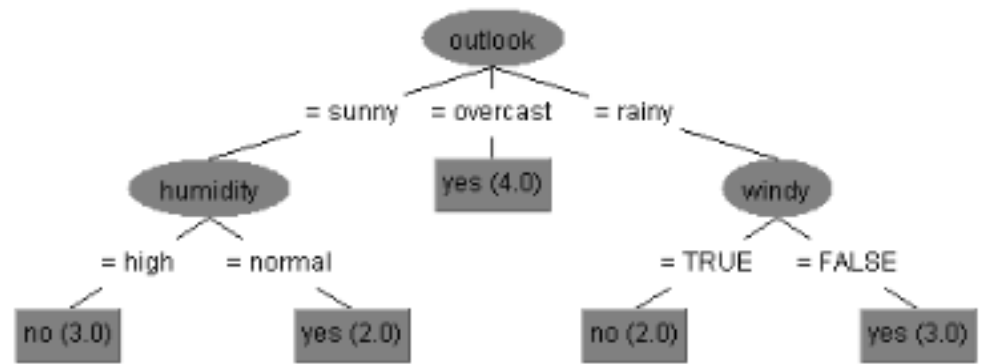
Diferentes Estruturas

Complexidade



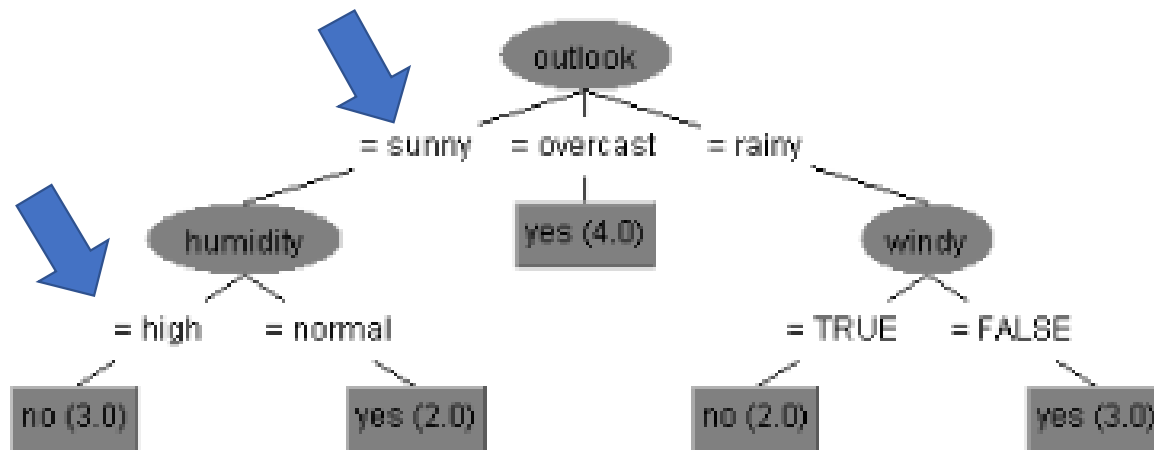
- **Métricas de Complexidade:**
- **Profundidade:** números de nós da raiz até as folhas
- **Largura:** número de nós em cada nível

Discretos vs Contínuos



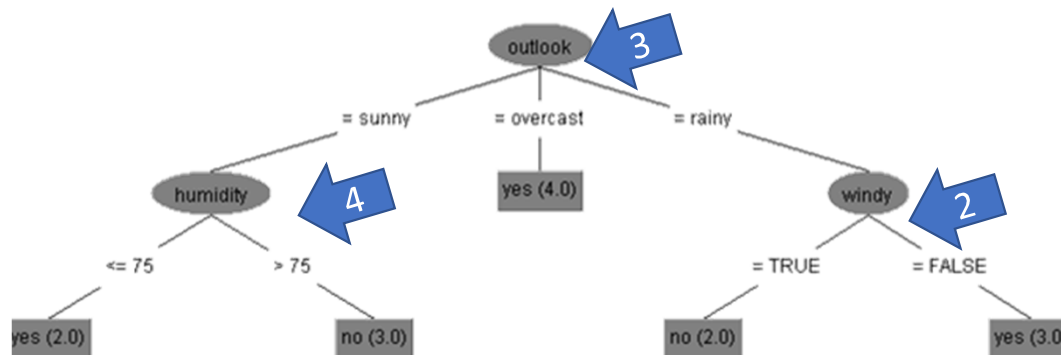
Processo de Classificação

1: outlook	2: temperature	3: humidity	4: windy
Nominal	Nominal	Nominal	Nominal
sunny	hot	high	FALSE



Indução da Árvore

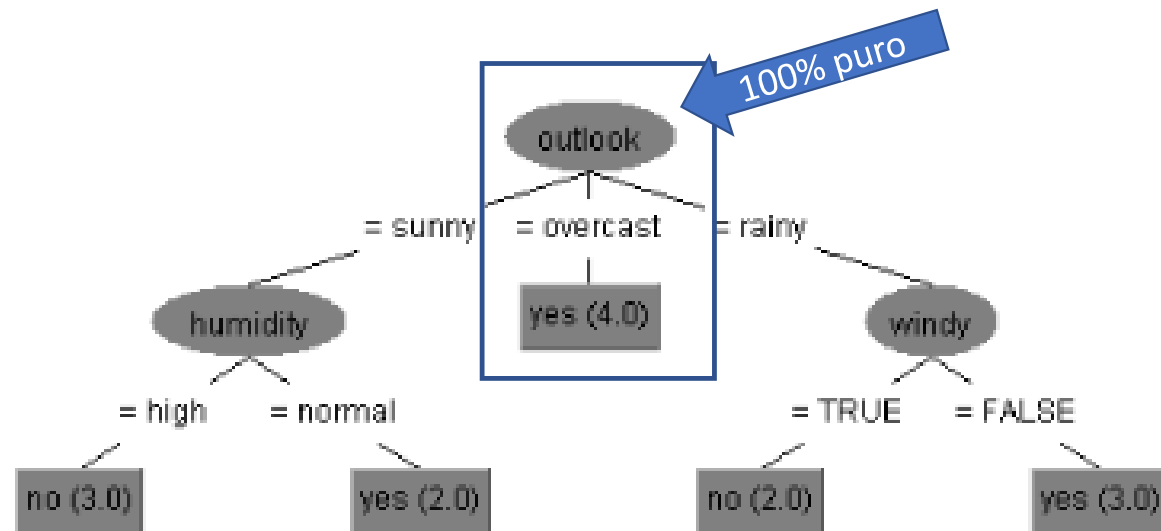
- Divisão:
 1. Em duas partes, ou em n partes
 2. Binários: divisão dupla
 3. Nominal: múltiplas divisões ou agrupando atributos em sub conjuntos
 4. Contínuo: comparação de valores ou discretização (duas divisões ou múltiplas divisões)



Divisão

➤ Objetivo é criar divisões o mais “puras” possíveis através de uma medida de pureza

- Gini
- Entropia
- Erro de classific



Condição de Parada

- Quando se chega a classe pura
- Número mínimo de observações em um nó
- A última partição não aumento a métrica de pureza

Poda

- Processo de reduzir o tamanho da árvore após a indução
- Árvores muito grandes estão sujeitas a super ajuste
- Mecanismos:
 - Checar se pares de nós podem ser fundidos aumentando a pureza
 - Checar se partições aumentaram a pureza

Prática

ID	CHEQUEESPECIAL	HISTORICO_CREDITO	RESIDENCIA	PAGADOR
1	<0	crítica/outr.exist.	própria	bom
2	0<=X<200	crítica/outr.exist.	própria	ruim
3	sem cheque	crítica/outr.exist.	própria	bom
4	<0	existente e pago	gratuitamente	bom
5	<0	atrasos anteriores	gratuitamente	ruim
6	sem cheque	existente e pago	gratuitamente	bom
7	sem cheque	existente e pago	própria	bom
8	0<=X<200	existente e pago	alugada	bom
9	sem cheque	existente e pago	própria	bom
10	0<=X<200	crítica/outr.exist.	própria	ruim
11	0<=X<200	existente e pago	alugada	ruim
12	<0	existente e pago	alugada	ruim
13	0<=X<200	existente e pago	própria	bom
14	<0	crítica/outr.exist.	própria	ruim
15	<0	existente e pago	alugada	bom