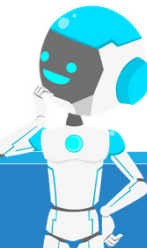
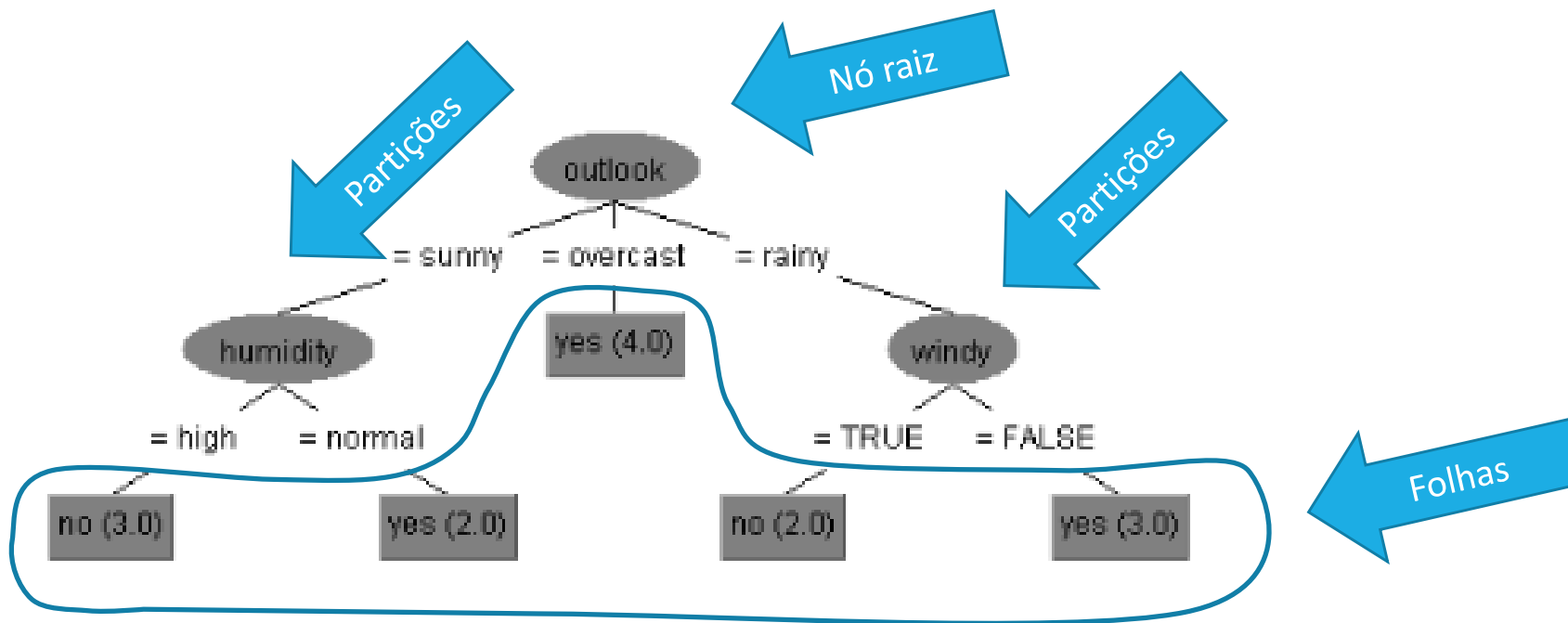
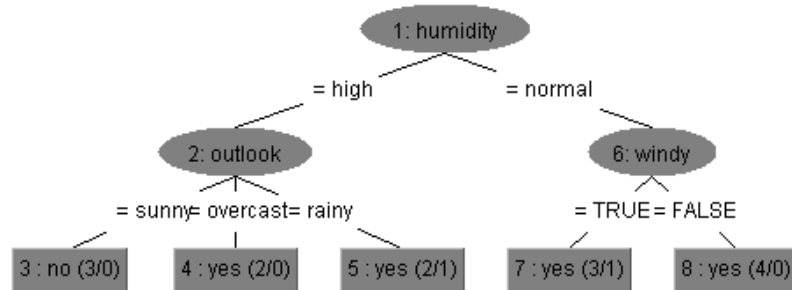
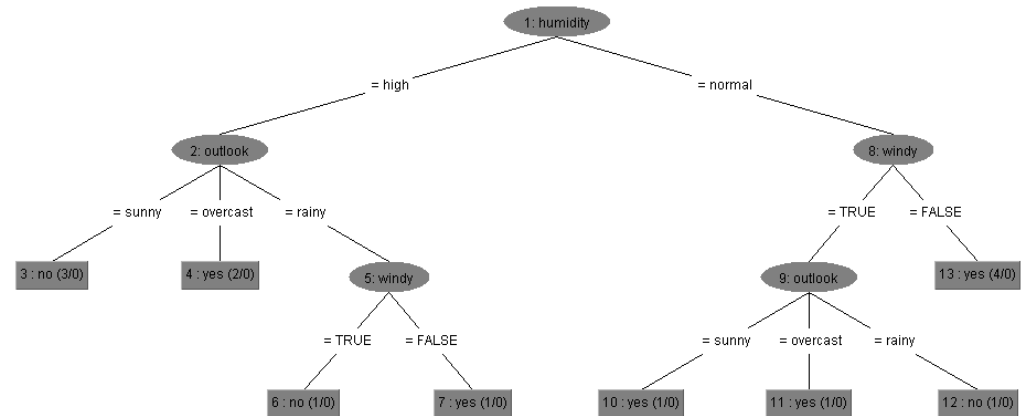
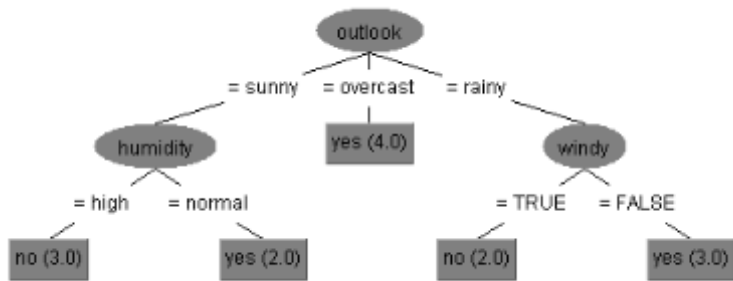


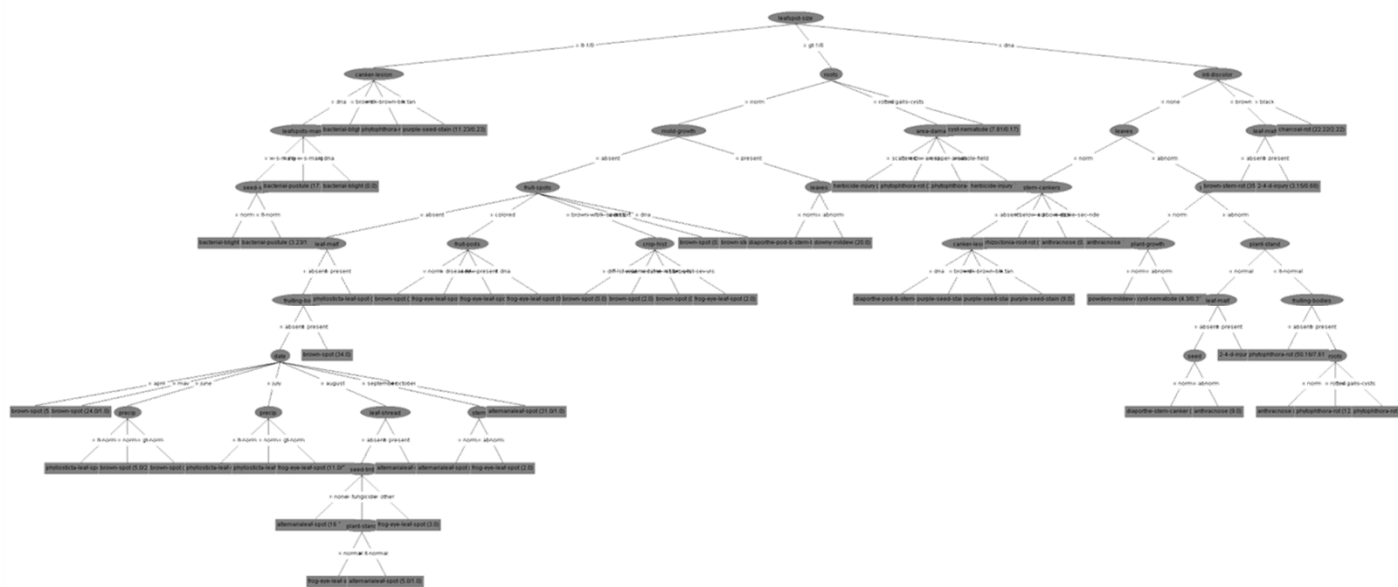
Estrutura



Diferentes Estruturas



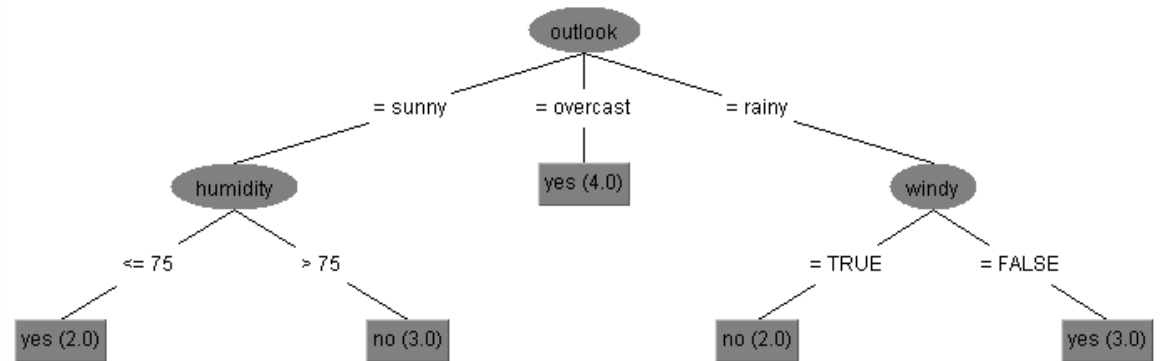
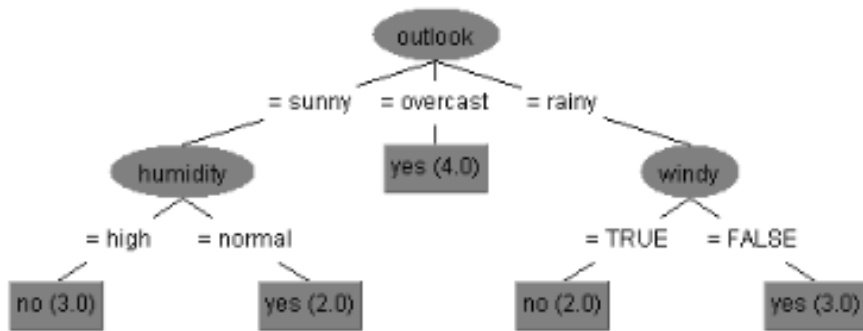
Complexidade



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

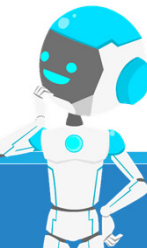
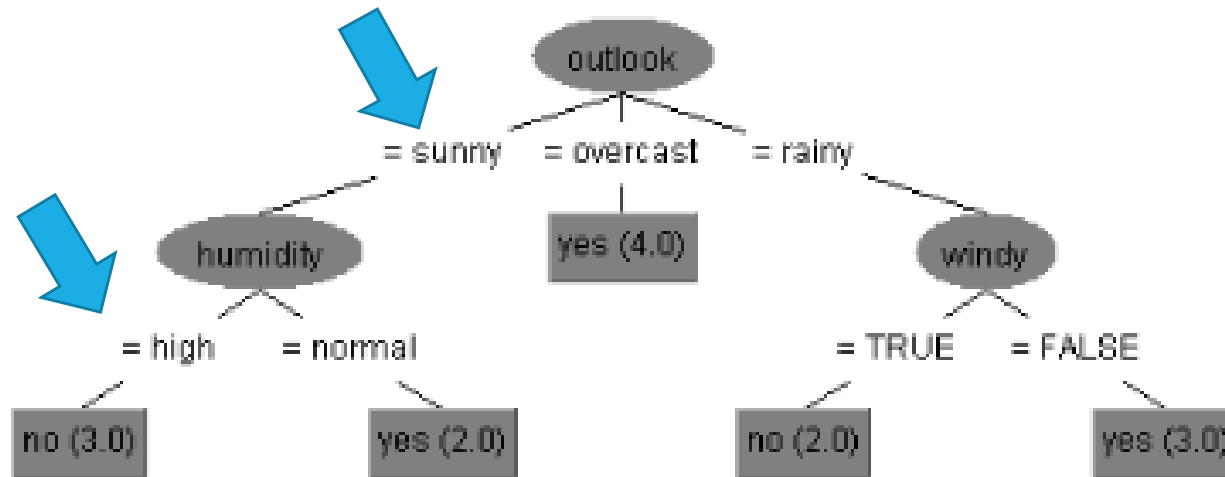


Discretos vs Contínuos



Processo de Classificação

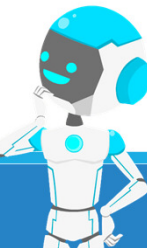
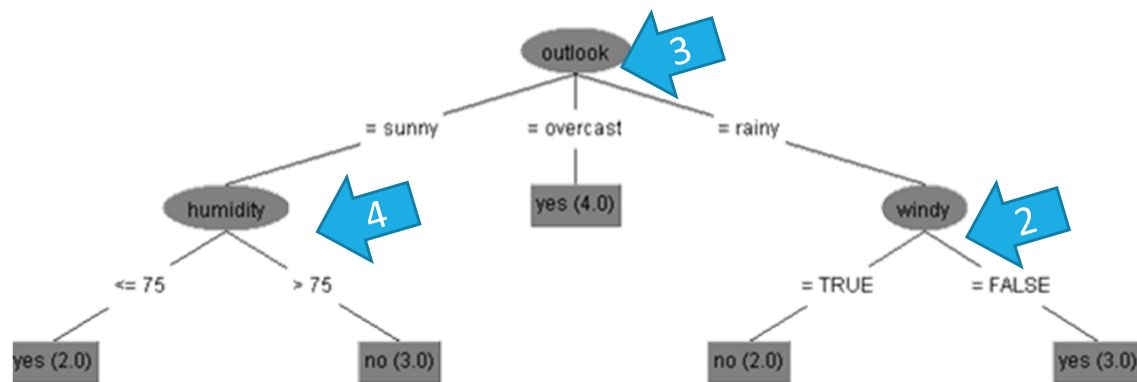
outlook	temperature	humidity	windy
sunny	hot	high	FALSE



Indução da Árvore

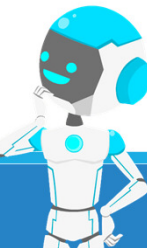
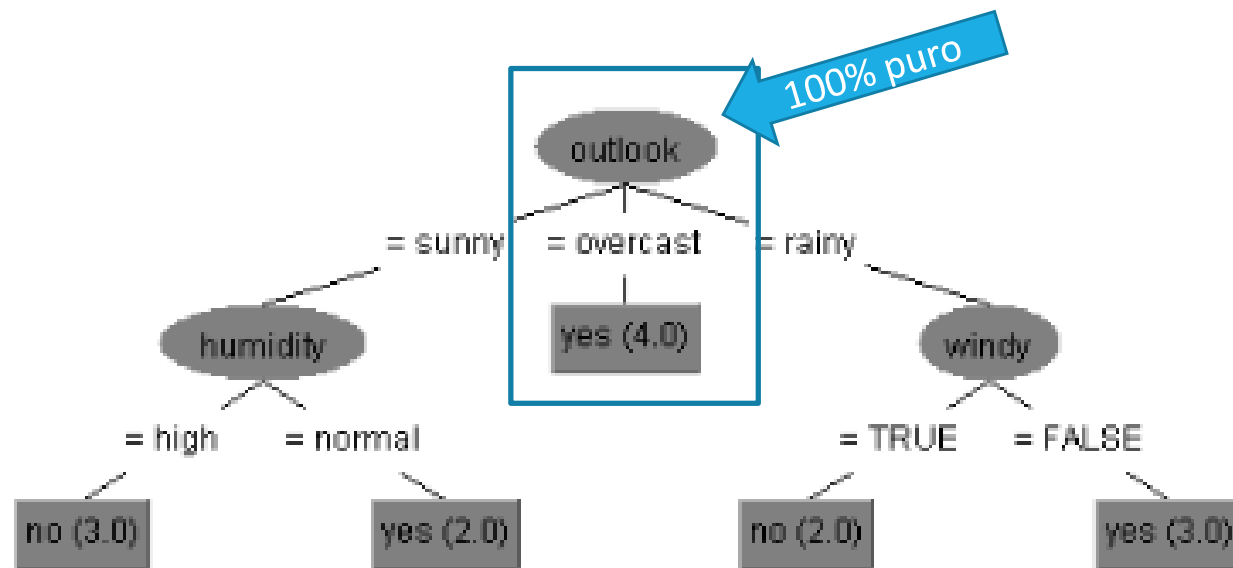
Divisão:

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 classificação



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