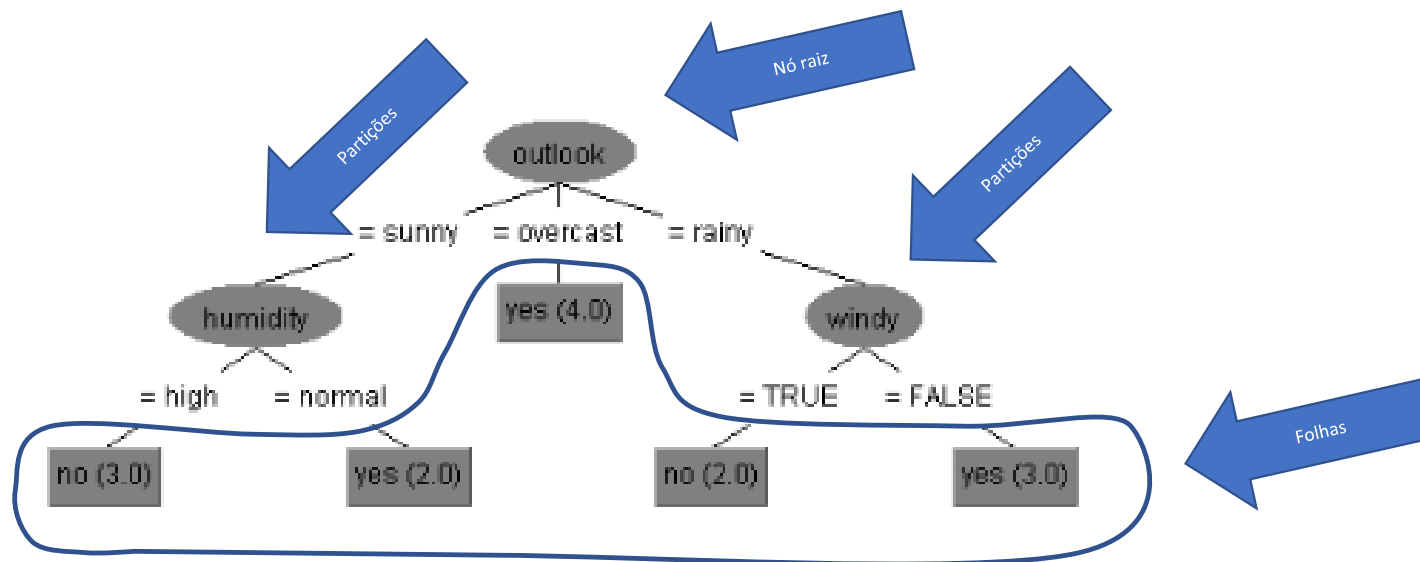


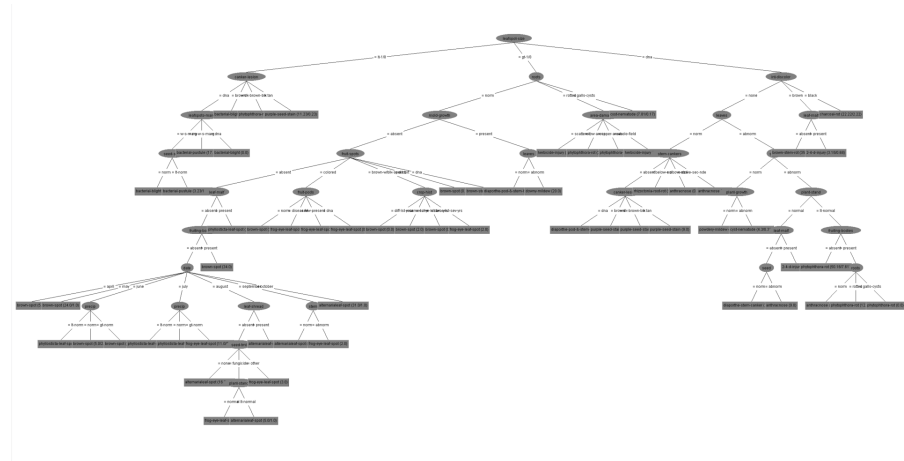
Árvore de Decisão



Complexidade

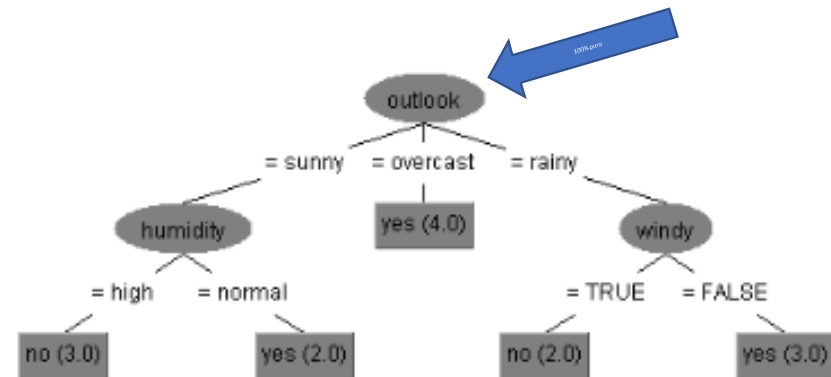
Métricas de Complexidade:

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



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

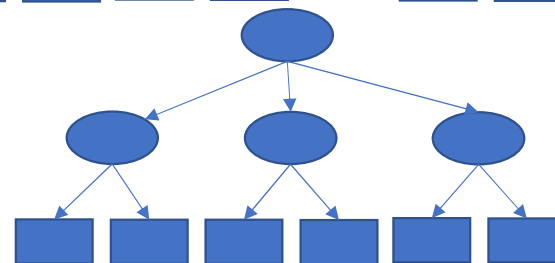
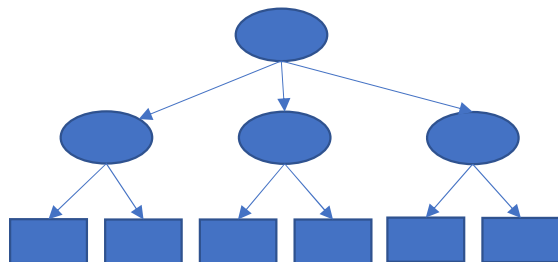
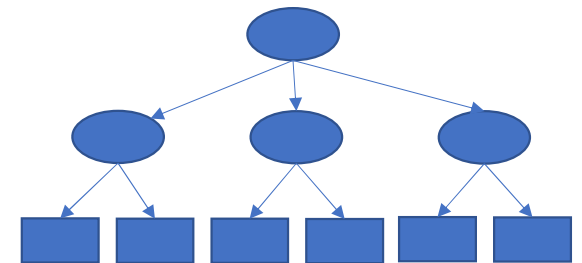
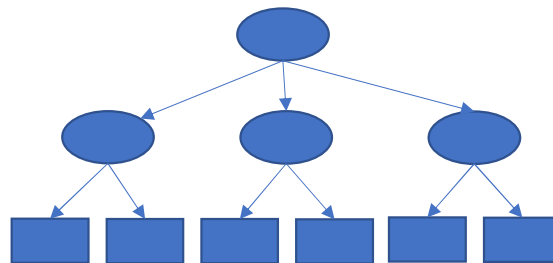
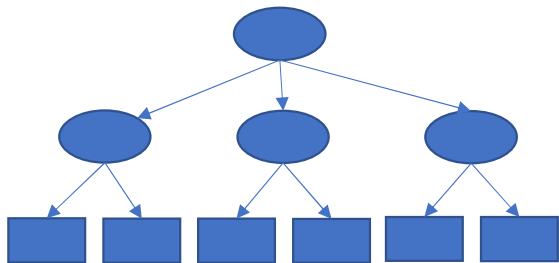
Random Forest

Induz diversas árvores de
decisão

Executa a processo de
classificação para cada árvore

Executa um processo de
votação para decisão da classe

Várias Árvores...

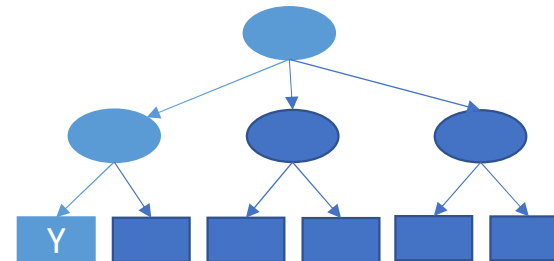
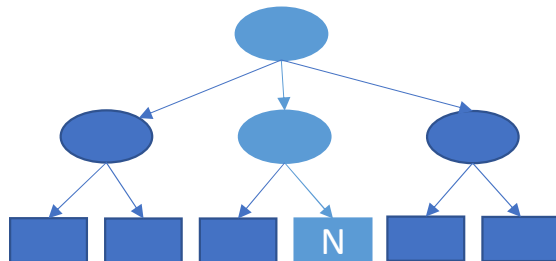
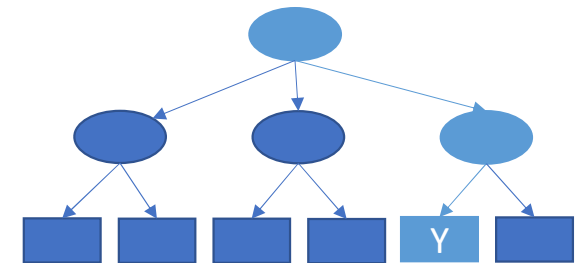
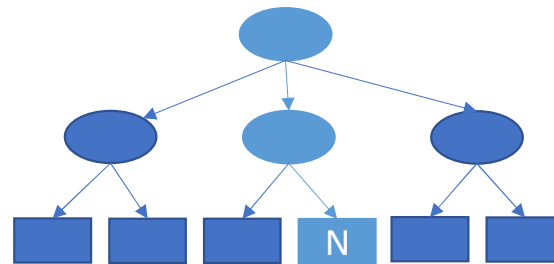
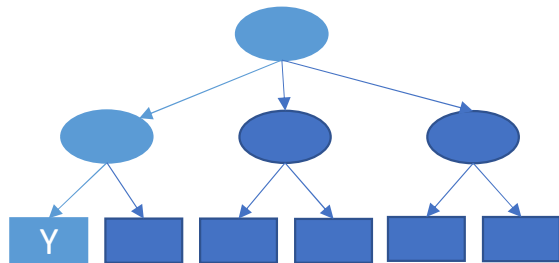


Múltiplas Árvores

Cria conjuntos de dados de treino de forma aleatória, porém com reposição (bootstrap)

Do total de atributos da relação, é selecionado um subconjunto de atributos aleatórios para cada árvore

Processo de Votação



Y = 3
N = 2



Hiper Parâmetros

- bootstrap: se deve ser usado na construção das árvores. (padrão True)
- maxBins: número máximo de valores na discretização de atributos contínuos. (padrão 32)
- maxDepth: profundidade máxima da árvore. (padrão 5)
- numTrees: número de árvores aleatórias que serão treinadas. (padrão 20)
- seed: semente para repetir o processo