



# Algoritmo de Random Forests ou Florestas Aleatórias

Gabriel Rosa e Vitor Oliveira



## Construindo o algoritmo

- Algoritmo bastante conhecido por sua simplicidade vinda da utilização de árvores de decisão e flexibilidade pela utilização de uma grande quantidade delas;
- Neste algoritmo são criadas várias árvores de decisão, a partir da escolha de dados para treinamento pelo método de bootstrapping, ou seja, dados aleatórios e escolhidos com repetição;
- Para definir a construção de cada árvore de decisão a ser utilizada são selecionados atributos aleatórios, como no exemplo a seguir;

# Exemplo de construção árvores de decisão no random forest

Dataset completo e correto

Dia	Aspecto	Temp.	Humidade	Vento	Jogar Ténis
D1	Sol	Quente	Elevada	Fraco	Não
D2	Sol	Quente	Elevada	Forte	Não
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não
D7	Nuvens	Fresco	Normal	Fraco	Sim
D8	Sol	Ameno	Elevada	Fraco	Não
D9	Sol	Fresco	Normal	Fraco	Sim
D10	Chuva	Ameno	Normal	Forte	Sim
D11	Sol	Ameno	Normal	Forte	Sim
D12	Nuvens	Ameno	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Ameno	Elevada	Forte	Não

1º Dataset para treinamento do modelo

Dia	Aspecto	Temp.	Humidade	Vento	Jogar Ténis
D1	Sol	Quente	Elevada	Fraco	Não
D2	Sol	Quente	Elevada	Forte	Não
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não
D7	Nuvens	Fresco	Normal	Fraco	Sim
D8	Sol	Ameno	Elevada	Fraco	Não
D9	Sol	Fresco	Normal	Fraco	Sim
D10	Chuva	Ameno	Normal	Forte	Sim
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não

# Construção das árvores

Dia	Aspecto	Temp.	Humidade	Vento	Jogar Ténis
D1	Sol	Quente	Elevada	Fraco	Não
D2	Sol	Quente	Elevada	Forte	Não
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não
D7	Nuvens	Fresco	Normal	Fraco	Sim
D8	Sol	Ameno	Elevada	Fraco	Não
D9	Sol	Fresco	Normal	Fraco	Sim
D10	Chuva	Ameno	Normal	Forte	Sim
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não

Escolha entre 2  
aleatórios

Aspecto ou  
Vento

Vento

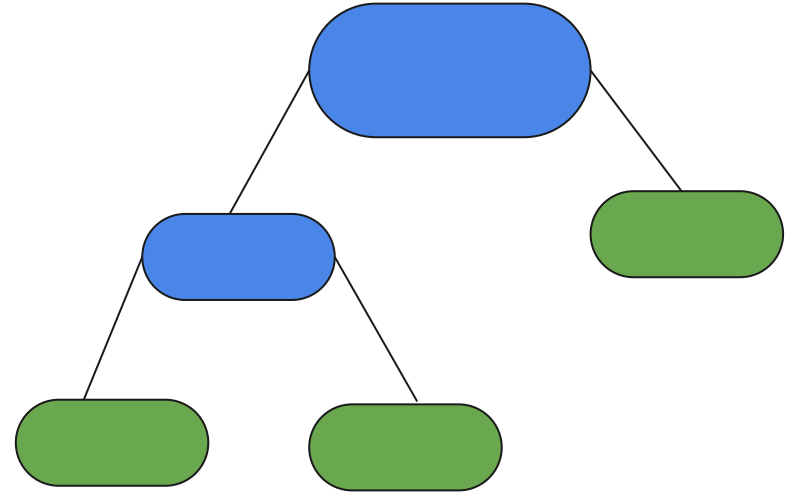
Outros 2  
atributos

Outros 2  
atributos

...

# Construção das árvores

Dia	Aspecto	Temp.	Humidade	Vento	Jogar Ténis
D1	Sol	Quente	Elevada	Fraco	Não
D2	Sol	Quente	Elevada	Forte	Não
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não
D7	Nuvens	Fresco	Normal	Fraco	Sim
D8	Sol	Ameno	Elevada	Fraco	Não
D9	Sol	Fresco	Normal	Fraco	Sim
D10	Chuva	Ameno	Normal	Forte	Sim
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não

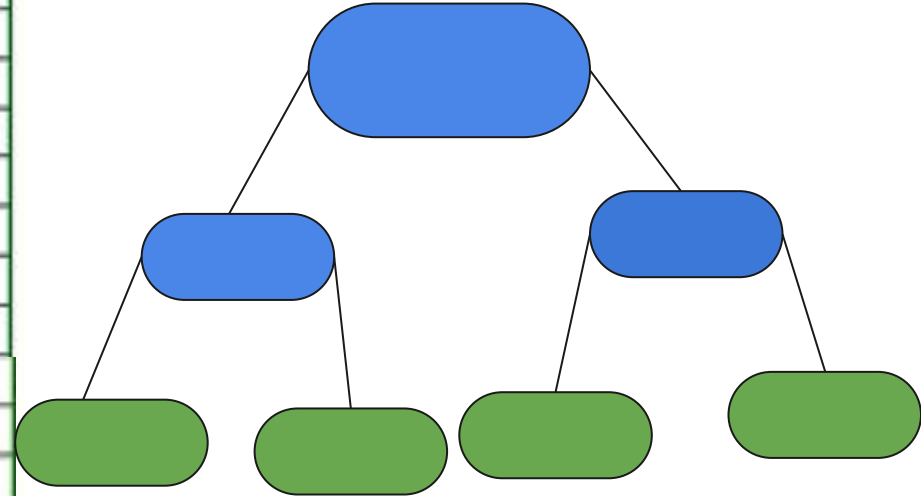




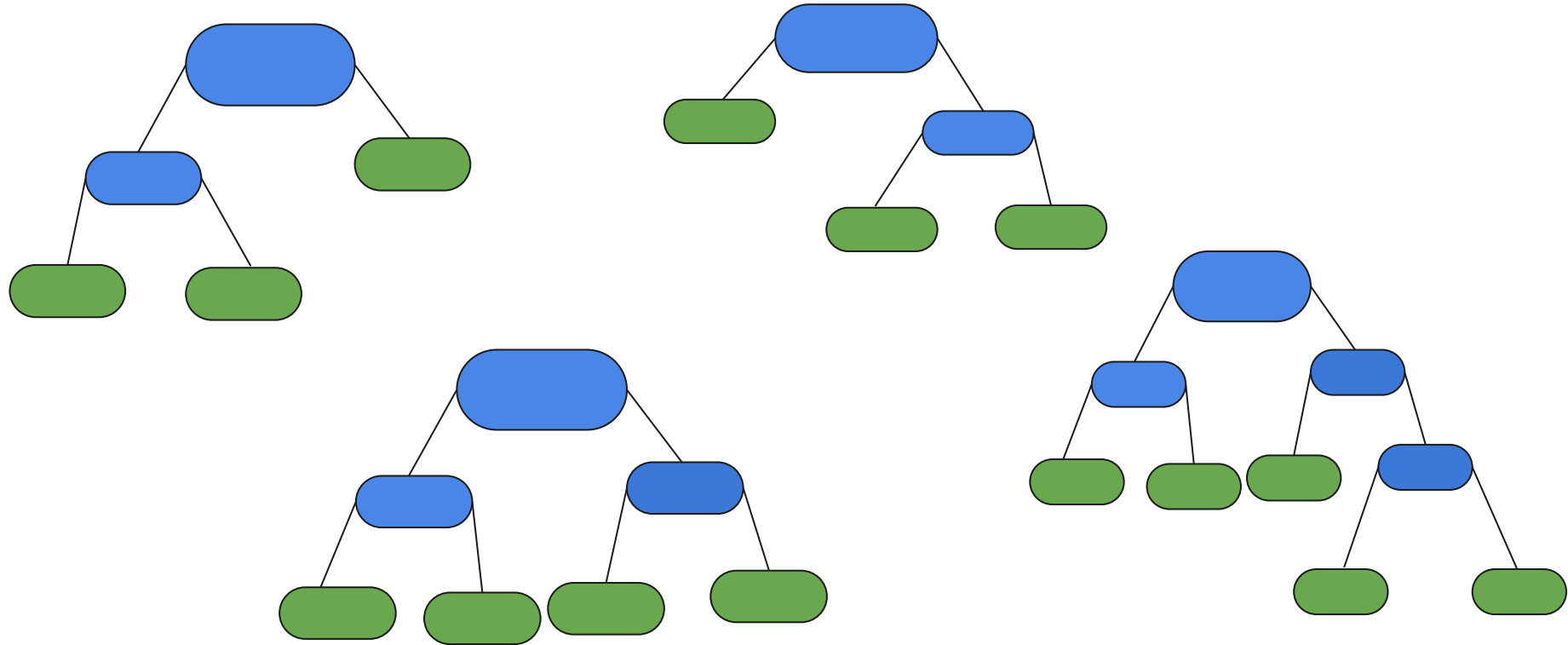
# Construção das árvores

Dia	Aspecto	Temp.	Humidade	Vento	Jogar Ténis
D1	Sol	Quente	Elevada	Fraco	Não
D2	Sol	Quente	Elevada	Forte	Não
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não
D7	Nuvens	Fresco	Normal	Fraco	Sim
D8	Sol	Ameno	Elevada	Fraco	Não
D9	Sol	Fresco	Normal	Fraco	Sim
D10	Chuva	Ameno	Normal	Forte	Sim
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não

Escolhendo outros atributos de maneira aleatória é possível chegar a outros tipos de árvores como:



# Construção das Árvores



# Após a definição das árvores

- Após o algoritmo definir a quantidade desejada de árvores na floresta, será possível conter uma grande variedade de diferentes árvores binárias para encontrar a previsão desejada dos novos dados a serem inseridos no modelo.
- A testagem do modelo acontece com os dados que ficaram de fora do set de treinamento, os quais forma o “Out-Of-Boot Dataset”, esses dados irão formar o nosso “Testing Data” que irá medir a precisão do nosso modelo.
- No exemplo mostrado, os dados que serão utilizados para teste serão os dados referentes aos dias 11, 12, 13 e 14.

D11	Sol	Ameno	Normal	Forte	Sim
D12	Nuvens	Ameno	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Ameno	Elevada	Forte	Não



# Avaliando o Modelo

A Partir desses nossos dados do nosso “Out-Of-Bag Dataset” iremos avaliar a performance do nosso modelo:

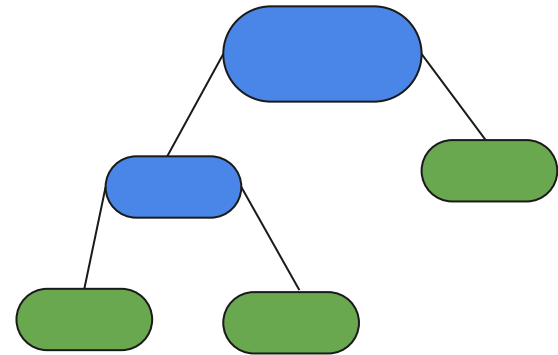
D11	Sol	Ameno	Normal	Forte	Sim
D12	Nuvens	Ameno	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Ameno	Elevada	Forte	Não

D11: SIM

D12: SIM

D13: NÃO

D14: NÃO



# Avaliando o Modelo

A Partir desses nossos dados do nosso “Out-Of-Bag Dataset” iremos avaliar a performance do nosso modelo:

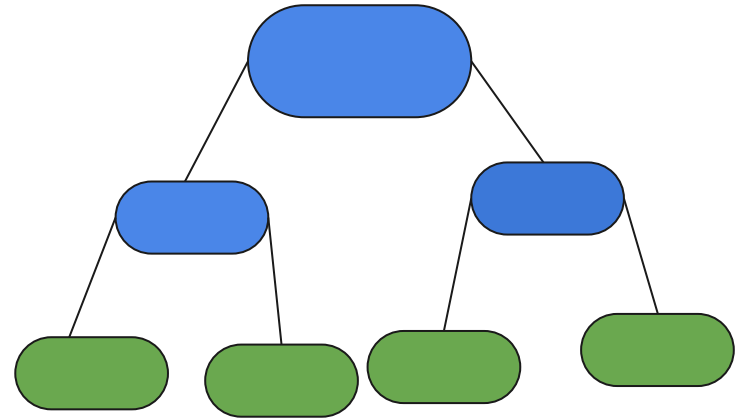
D11	Sol	Ameno	Normal	Forte	Sim
D12	Nuvens	Ameno	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Ameno	Elevada	Forte	Não

D11: NÃO

D12: SIM

D13: SIM

D14: SIM



# Avaliando o Modelo

A Partir desses nossos dados do nosso “Out-Of-Bag Dataset” iremos avaliar a performance do nosso modelo:

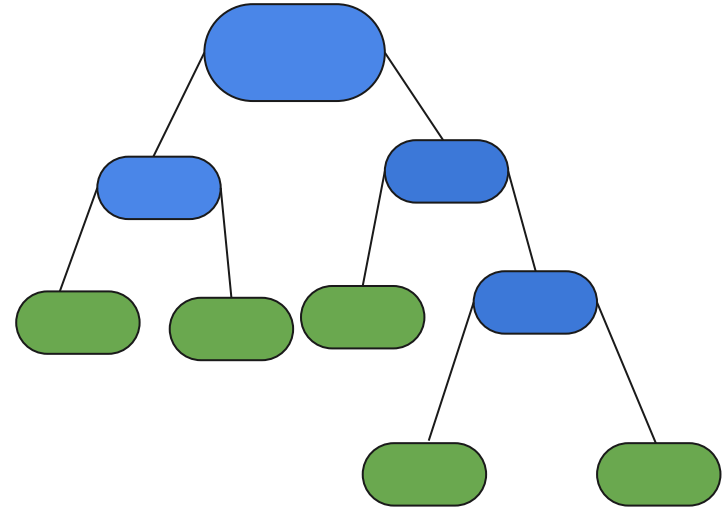
D11	Sol	Ameno	Normal	Forte	Sim
D12	Nuvens	Ameno	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Ameno	Elevada	Forte	Não

D11: SIM

D12: SIM

D13: SIM

D14: SIM

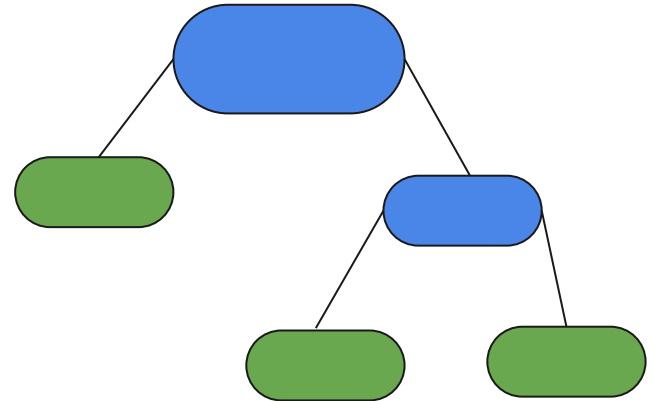


# Avaliando o Modelo

A Partir desses nossos dados do nosso “Out-Of-Bag Dataset” iremos avaliar a performance do nosso modelo:

D11	Sol	Ameno	Normal	Forte	Sim
D12	Nuvens	Ameno	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Ameno	Elevada	Forte	Não

D11: SIM  
D12: NÃO  
D13: SIM  
D14: SIM



# Avaliando o Modelo

Com a passagem pelas 4 árvores criadas, medimos o quão bom nosso modelo de fato é:

D11	Sol	Ameno	Normal	Forte	Sim
D12	Nuvens	Ameno	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Ameno	Elevada	Forte	Não

Dados os resultados tivemos que, de acordo com os resultados, o Random Forest decide de acordo com a maioria, logo os resultados previstos para os dias 11, 12, 13 e 14 seriam:

D11 : SIM

D12: SIM

D13: SIM

D14: SIM ( Erro na previsão)

Logo temos uma acurácia de 3/4 nesse algoritmo, para esse Dataset

D11:

SIM / NÃO / SIM / SIM = 3 Corretos e 1 Falso

D12:

SIM / SIM / SIM / NÃO = 3 Corretos e 1 Falso

D13:

NÃO / SIM / SIM / SIM = 3 Corretos e 1 Falso

D14:

NÃO / SIM / SIM / SIM = 1 Corretos e 3 Falsos

# Avaliando o Modelo

Após a avaliação do nosso modelo, podemos tentar refina-lo, uma maneira seria o aumento de opções para decisão de raízes na construção das árvores:

Dia	Aspecto	Temp.	Humidade	Vento	Jogar Tênis
D1	Sol	Quente	Elevada	Fraco	Não
D2	Sol	Quente	Elevada	Forte	Não
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não
D7	Nuvens	Fresco	Normal	Fraco	Sim
D8	Sol	Ameno	Elevada	Fraco	Não
D9	Sol	Fresco	Normal	Fraco	Sim
D10	Chuva	Ameno	Normal	Forte	Sim
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não

Escolha entre  
3 aleatórios

Aspecto, Vento  
ou Temp

Vento

Outros 3  
atributos

Outros 3  
atributos





## Concluindo

- O Random forest consegue utilizar da simplicidade das árvores de decisão para construção de seu algoritmo ao mesmo tempo que consegue se adaptar e generalizar melhor dados novos ao utilizar de várias árvores diferentes, construídas por combinações com repetição do dataset original.
- Porém como desvantagens ele acaba sendo um pouco mais demorado que outros algoritmos devido a essa construção de variadas árvores binárias além da sua capacidade de overfitting vinda da utilização das árvores.



# Demonstração do Random Forest no WEKA

