



# Random Forest

Aprendizado de Máquina

[elias.rodriques@paulista.ifpe.edu.br](mailto:elias.rodriques@paulista.ifpe.edu.br)



# Floresta Aleatória (Random Forest)

- Algoritmo de aprendizagem de máquina
  - Flexível
  - Fácil de usar
  - Produz excelentes resultados
- É também um dos algoritmos mais utilizados
  - Devido à sua simplicidade
  - Pode ser utilizado para tarefas de:
    - Classificação
    - Regressão
- Extensão das Árvore de Decisão
  - Modelo ensemble



# Floresta Aleatória (Random Forest)

## Árvore de Decisão

- Baseadas em dados
- Constrói uma única árvore de decisão
  - Um modelo simples do todo
- Muitas vezes o modelo é:
  - Muito simples
  - Muito específico
  - Baixo custo computacional
- Alta susceptibilidade a:
  - Overfitting
  - Ruídos nos dados

## Random Forest

- Baseadas em dados
- Constrói Várias árvores de decisão
  - Múltiplos modelos de partes do todo
- O conjunto de modelos é:
  - Mais complexo
  - Mais geral
  - Alto custo computacional
- Pouca susceptibilidade a:
  - Overfitting
  - Ruídos nos dados



# Random Forest

“O algoritmo de florestas aleatórias cria várias árvores de decisão e as combina para obter uma predição com maior acurácia e mais estável”



# Floresta Aleatória (Random Forest)

Adiciona aleatoriedade extra ao modelo

## Árvore de Decisão

- Todo conjunto de treinamento
- Na partição de nodos:
  - Busca pela melhor característica global
    - Ex: Entropia
  - Repete o procedimento
    - Excluindo as características já selecionadas

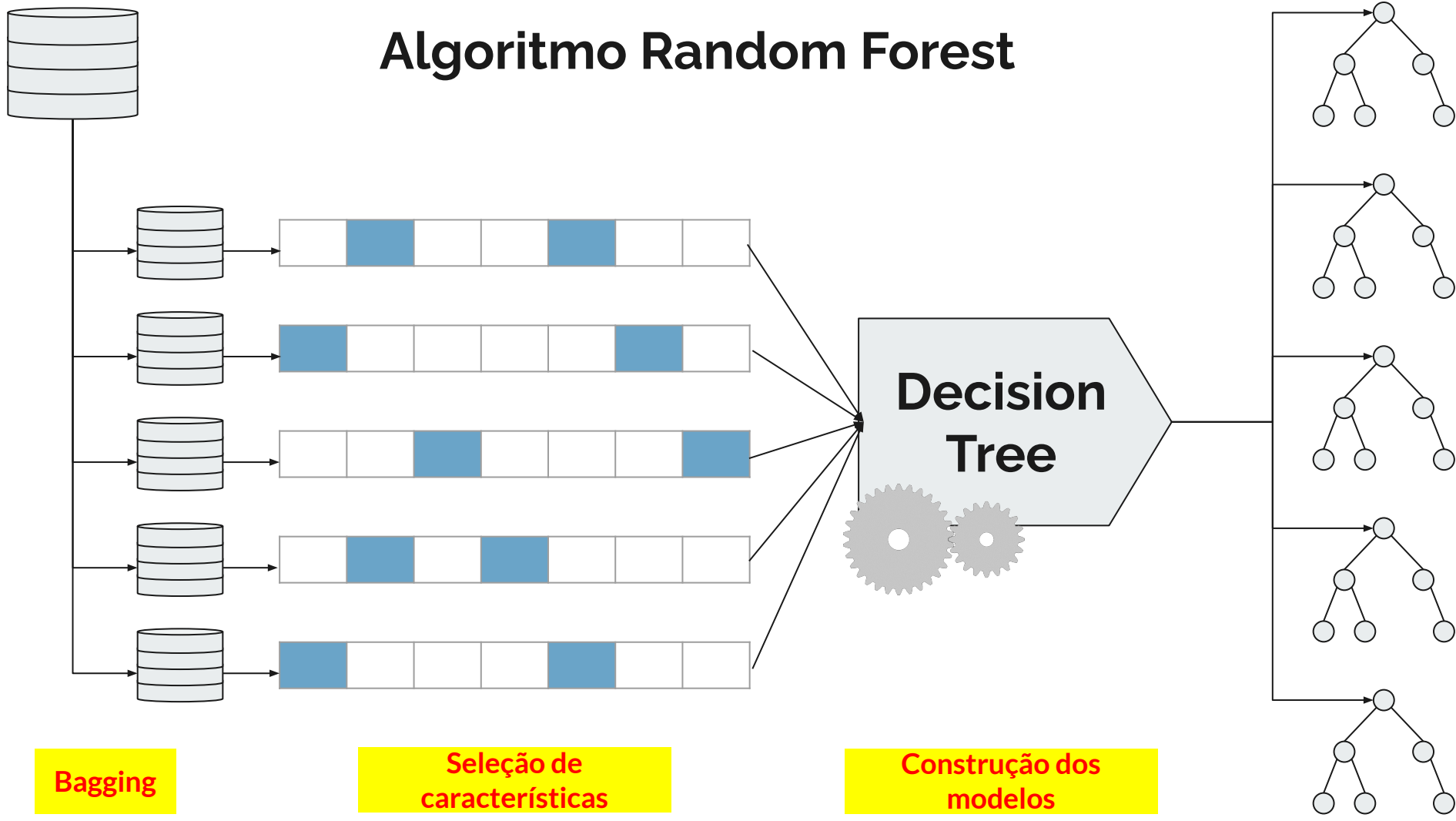
Leva a geração de modelos melhores

## Random Forest

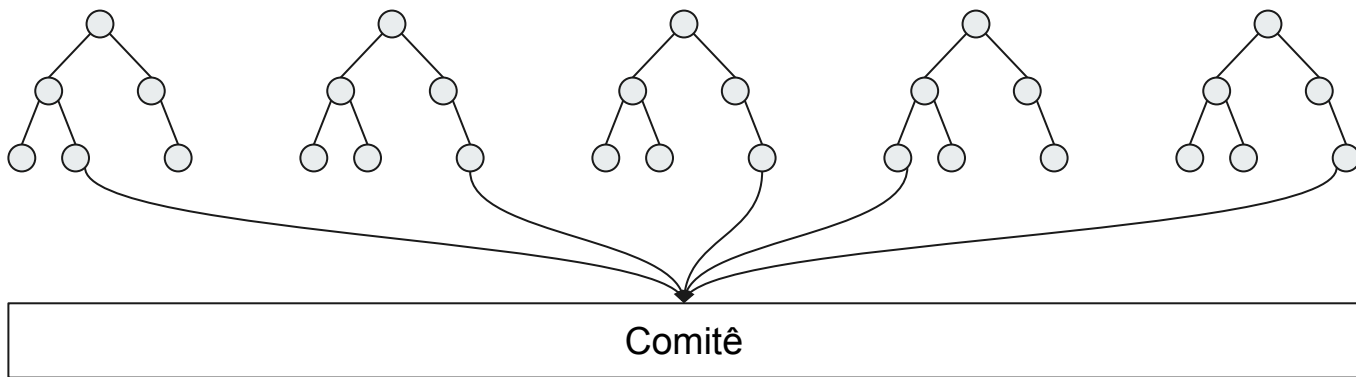
- Bagging do conjunto de treinamento
- Na partição de nodos:
  - Busca a melhor característica em um subconjunto aleatório das características
    - Ex: Aleatoriedade + entropia
  - Repete o procedimento
    - Aleatoriedade no:
      - Subconjunto de treino
      - Seleção das características

Cria uma grande diversidade

# Algoritmo Random Forest



## Decisão do Random Forest



### Resultado Final:

**Regressão** pode-se utilizar a média dos valores

### Resultado Final:

**Classificação** pode-se utilizar a moda dos valores



# Random Forest no SKLearn

## Regressão

[sklearn.ensemble.RandomForestRegressor](#)

## Classificação

[sklearn.ensemble.RandomForestClassifier](#)



# Recesso:

**31/12/2021 à 09/01/2022**

Prova: 10/01/2022

Segunda chamada: 17/01/2022

Prova Final: 24/01/2022

---