

CURSO SUPERIOR DE TECNOLOGIA EM CIÊNCIA DE
DADOS - UEPB

INTRODUÇÃO AO MACHINE LEARNING COM APLICAÇÃO EM R

MSC. VIVIANE COSTA SILVA

**DOUTORANDA DO PROGRAMA DE PÓS-GRADUAÇÃO EM
ESTATÍSTICA
E EXPERIMENTAÇÃO AGROPECUÁRIA - UFLA**

ORIENTADOR: PROFº DR. LUIZ RICARDO NAKAMURA



MONARD, M. C.

O APRENDIZADO DE MÁQUINA É UM RAMO DA INTELIGÊNCIA ARTIFICIAL (IA) DEDICADO AO DESENVOLVIMENTO DE SISTEMAS E ALGORITMOS CAPAZES DE APRENDER E DE MELHORAR SEU DESEMPENHO DE MANEIRA AUTOMÁTICA, SEM A NECESSIDADE DE INTERFERÊNCIA HUMANA.

APRENDIZADO DE MÁQUINAS

Aprendizado é qualquer processo pelo qual um sistema melhora sua performance pela experiência.

• *Herbert Alexander Simon*



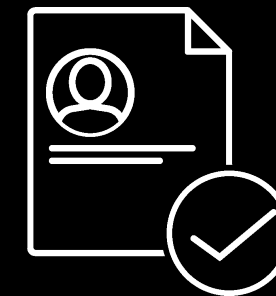
IMPORTÂNCIA



Automação e
Eficiência



Tomada de
Decisões
Baseada em
Dados



Avanços
Científicos e
Inovação

PORQUE TRABALHAR COM ML

Inundação de
dados
disponíveis

Incremento de
força
computacional

Progresso crescente de
algoritmos disponíveis e
teoria desenvolvida por
pesquisadores

Aumento no
suporte/apoio
das indústrias

TIPOS DE

APRENDIZADO

01

Aprendizado
Supervisionado

02

Aprendizado não
Supervisionado

03

Aprendizado por
reforço

APRENDIZADO SUPERVISIONADO

- Classificação de e-mails
- Diagnóstico médico
- Reconhecimento de voz
- Tradução automática

EXEMPLOS

APRENDIZADO NÃO SUPERVISIONADO

- Clusterização de clientes
- Segmentação de mercado
- Análise de redes sociais
- Detecção de anomalias

EXEMPLOS

APRENDIZADO POR REFORÇO

- Jogos de vídeo
- Robótica
- Treinamento de agentes de jogos de estratégia
- Treinamento de robôs para manipulação de objetos

EXEMPLOS

- Divisão de Dados
- Técnicas de pré-processamento
- Cross Validation
- Otimização de hiperparâmetros
- Treinamento
- Métricas de desempenho
- Teste

ETAPAS DO MACHINE LEARNING

DIVISÃO DOS DADOS

**BANCO DE
DADOS**



TREINAMENTO

TESTE

TECNICAS DE **PROCESSAMENTO**

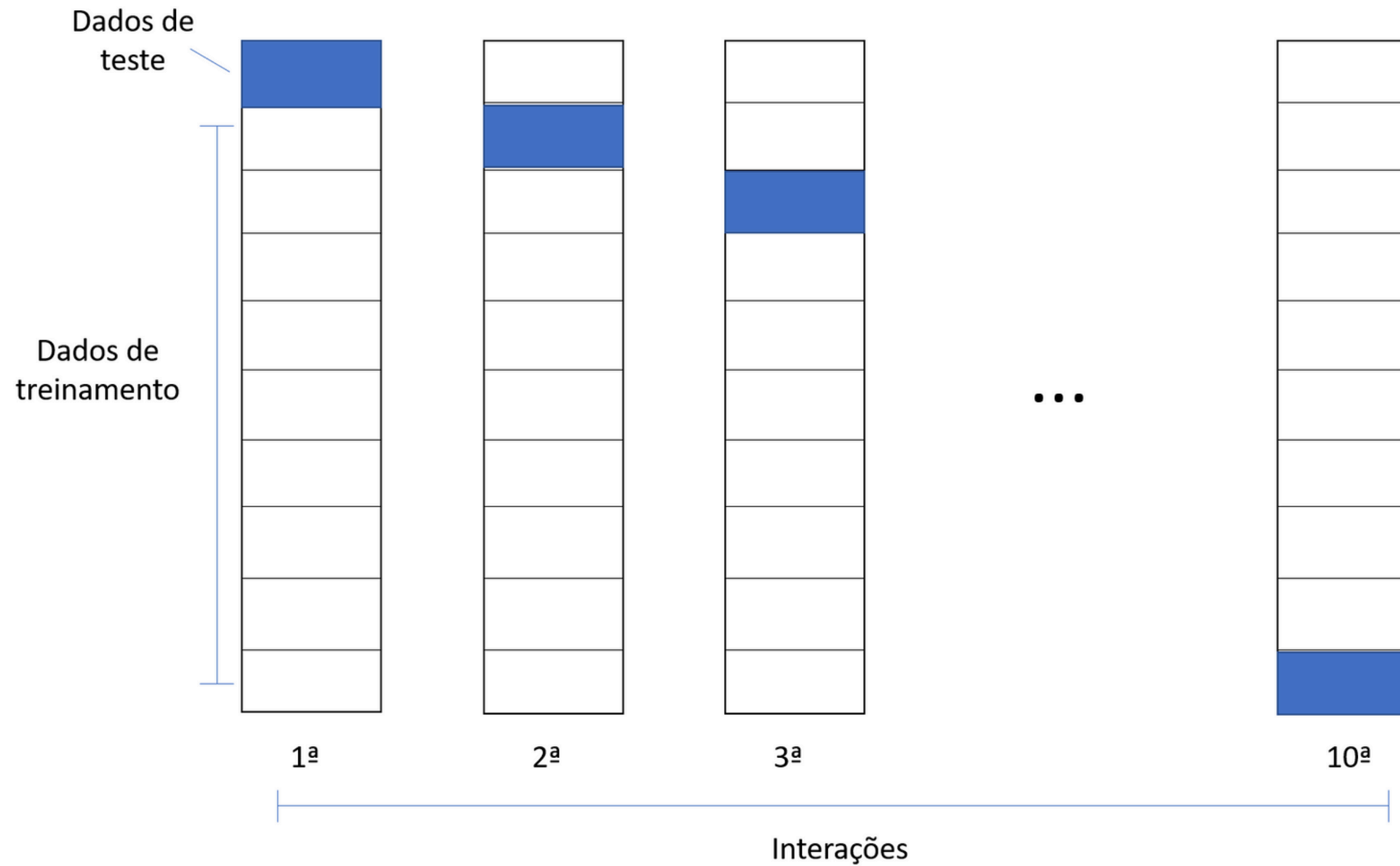
Feature selection

Imputação

Normalização

Redução de
dimensionalidade

CROSS VALIDATION



OTIMIZAÇÃO DOS HIPERPARÂMETROS

Grid Search

Random Search

Otimização bayesiana

MÉTRICAS DE DESEMPENHO

Análise supervisionada



```
graph TD; A[Análise supervisionada] --> B[Regressão]; A --> C[Classificação];
```

Regressão

MAE
RSME
RSE
MAPE
 R^2

Classificação

Acurácia;
Precisão;
Taxa de erro;
Recall (r);
Especificidade;
Curva ROC.

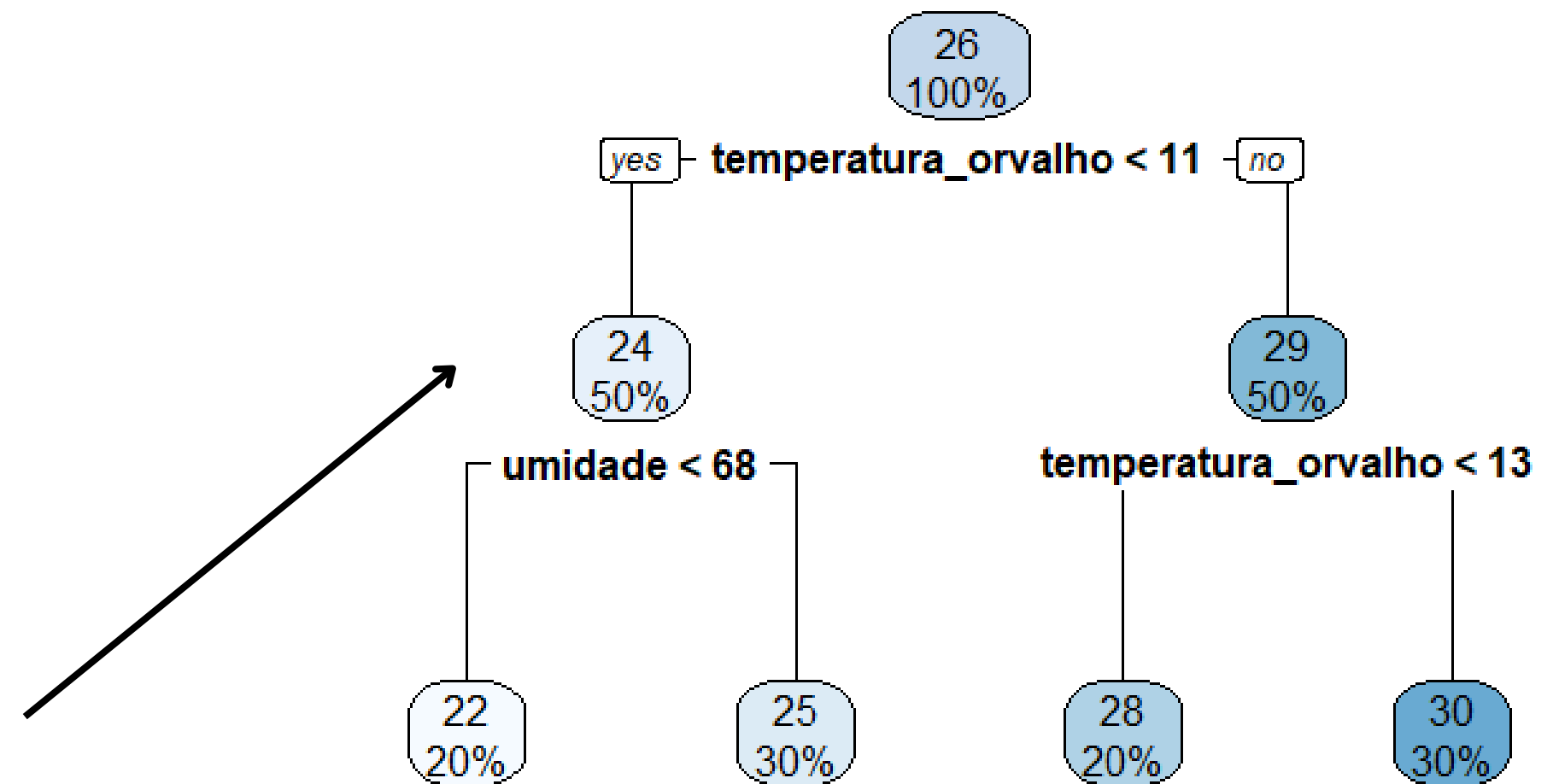
ESCOLHA DOS ALGORITMOS

DECISION TREES

BANCO DE DADOS



AMOSTRA



DECISION TREES

HIPERPARÂMETROS DO MODELO

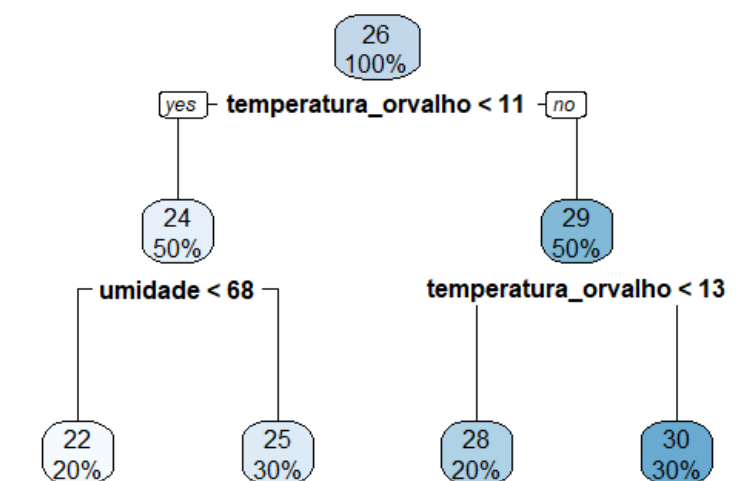
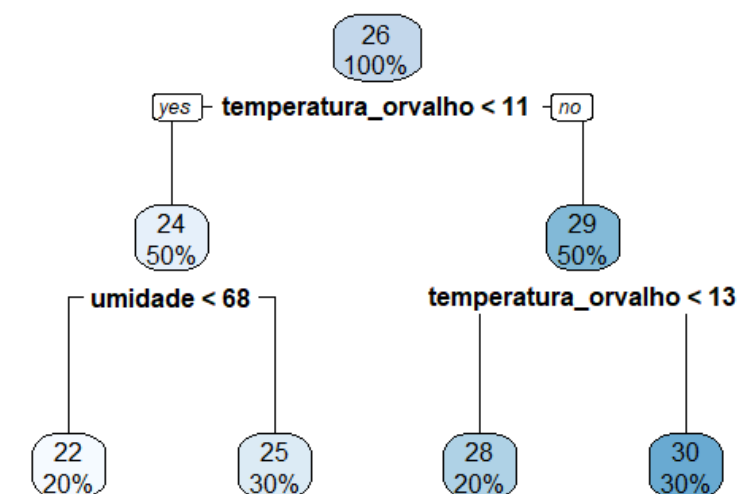
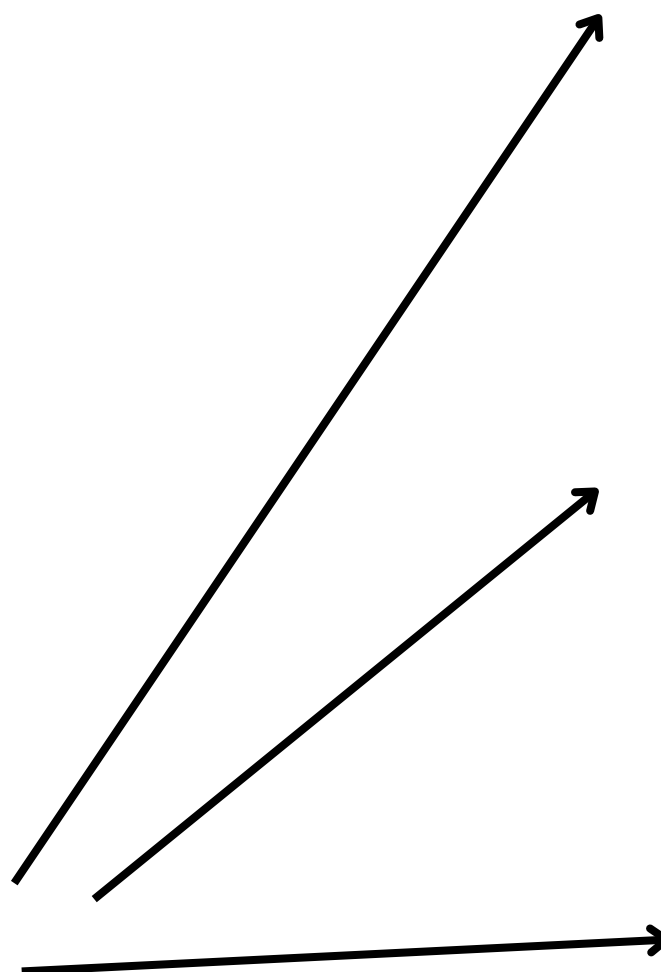
- Complexidade de Custo
- Número Mínimo de Amostras para Divisão

RANDOM FLOREST

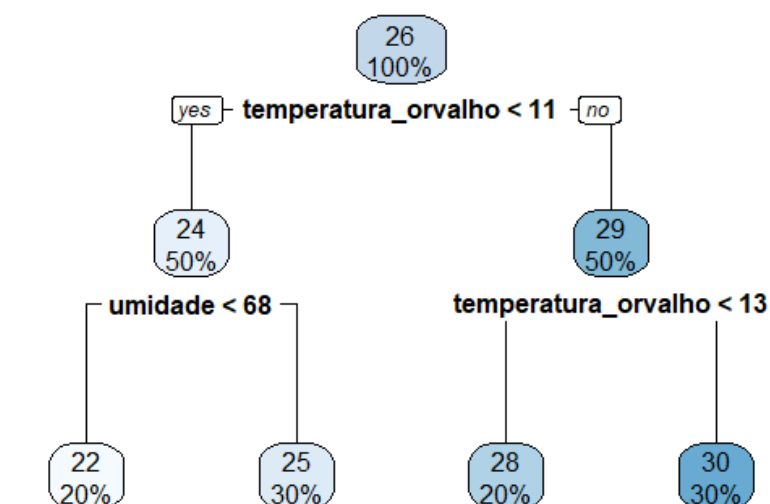
BANCO DE DADOS



...



...



AMOSTRAS

RANDOM FLOREST

HIPERPARÂMETROS DO MODELO

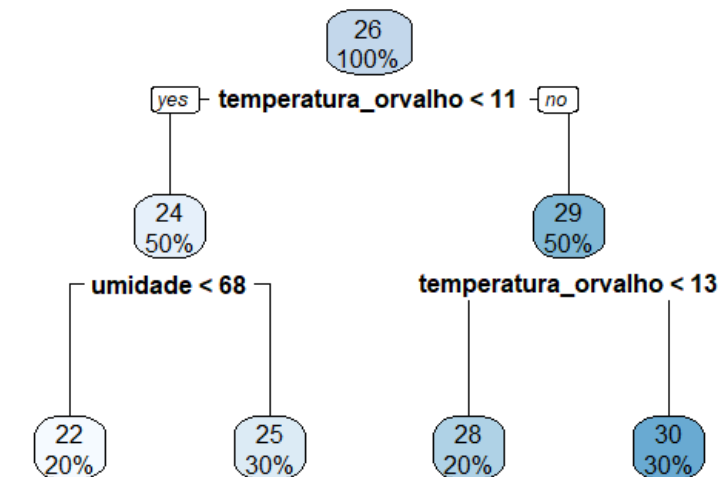
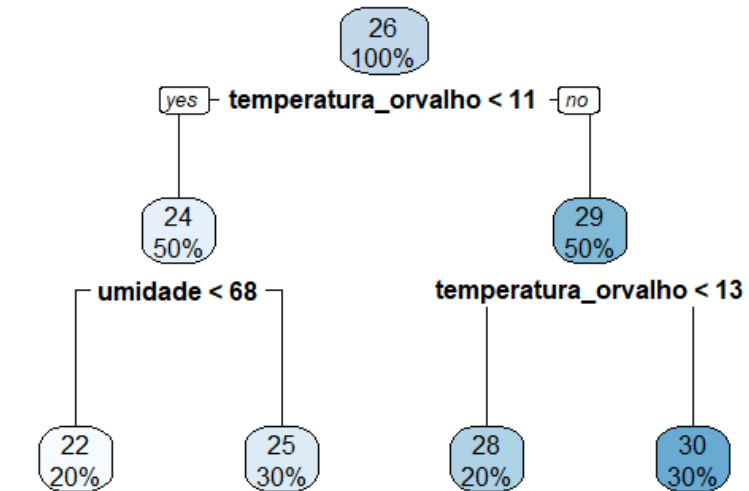
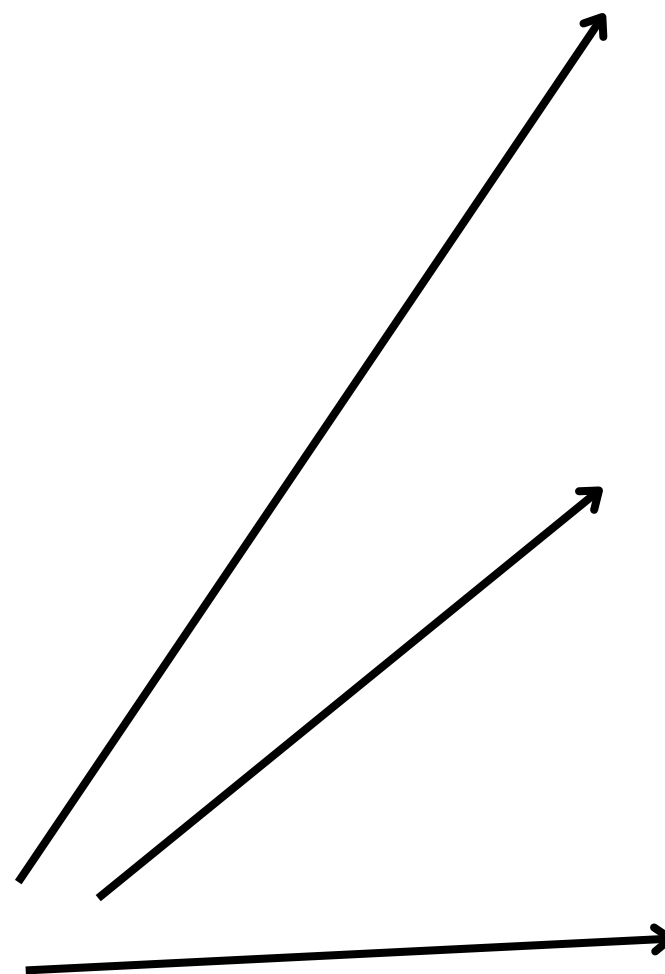
- Número máximo de atributos (características)
- Número Mínimo de Amostras para Divisão
- Número de árvores a serem construídas

XGBOOST

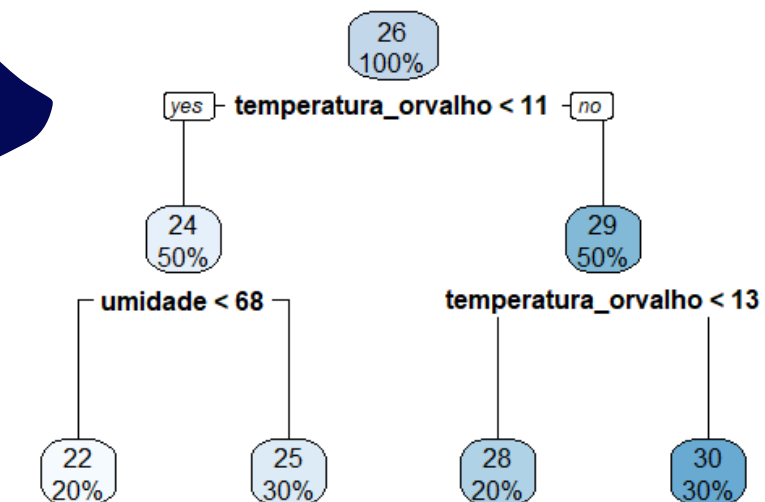
BANCO DE DADOS



...



...



AMOSTRAS

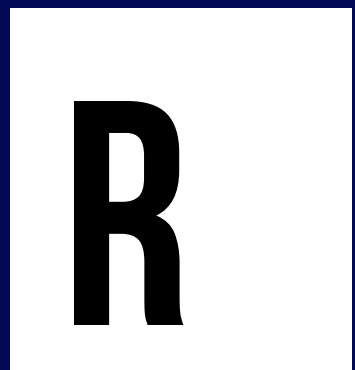
XGBOOST

HIPERPARÂMETROS DO MODELO

- Profundidade da árvore
- Taxa de aprendizado
- Número máximo de atributos (características)
- Número Mínimo de Amostras para Divisão
- Número de árvores a serem construídas

TREINAMENTO E TESTE

SOFWTARE

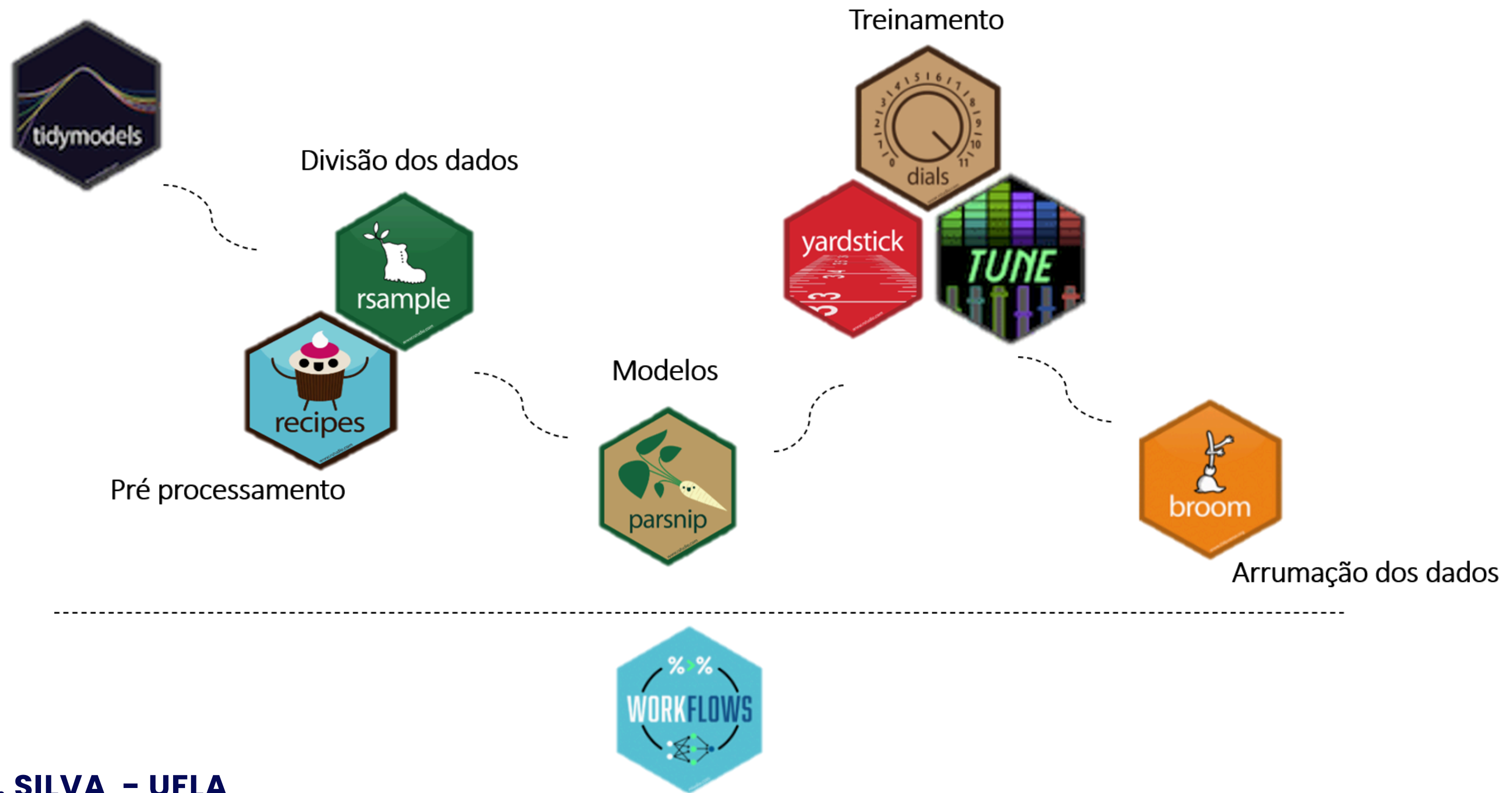


Linguagem de
programação e ambiente
de computação estatística

Código aberto e gratuito

Vasta gama de pacotes e
bibliotecas

O PACOTE TIDYMODELS



MAIS INFORMAÇÕES

■ Site do pacote

<https://www.tidymodels.org/>

PACOTE TIDYMODELS

MÃOS NA MASSA...

OU MELHOR!

MÃOS NO R!