



**PYCARET**

# **AutoML com Pycaret**

 **Stack**Academy

# Objetivo

- Entender o que é AutoML.
- Desenvolver **experimentos de modelagem** utilizando o Pycaret.
- Ter mais **produtividade e velocidade** no desenvolvimento de projetos de Machine Learning.
- Saber se **comparar modelos, técnicas, algoritmos e pipelines** para implementação de modelos de Machine Learning.
- Estar pronto para **enfrentar qualquer desafio**.



# Pra quem é esse curso?

- Cientistas de Dados.
- Engenheiros de Machine Learning.
- Gerentes e Heads de Tecnologia.



# Responsabilidades dos Profissionais

- ★ **Cientistas de Dados.**
  - Definir etapas para compor o Pipeline de Machine Learning.
    - Amostragem.
    - Preparação de dados.
      - Missing Values
      - Conversão de tipos de dados.
      - One Hot, Cardinal, Ordinal Encoding.
      - Desbalanceamento de classes.
    - Transformação.
      - Normalização.
      - Escala e Transformação.
    - Engenharia e Seleção de Features.
      - Interação de Features.
      - Agrupamento.
      - Binarização.
      - Importância de Features.
      - Remoção de Multicolinearidade.
      - Análise de componentes principais.
      - Ignorar baixa variância.
    - Definição de tarefas.
    - Otimização e Validação.
  - Definir técnicas e métodos a serem executados.
    - Algoritmos.
    - Estratégia para seleção de hiperparâmetros.
    - Métricas de avaliação.
  - Deploy, Monitoramento e análise.
    - Interpretação de Resultados.
    - Configurar experimentos e integrações.



# Responsabilidades dos Profissionais

- Engenheiro de Machine Learning
  - **Todas as atribuições do Cientista de Dados citadas anteriormente e ainda.**
    - Definir ambiente e arquitetura para treinamento e otimização de modelos.
      - On Premise, On Cloud
    - Implementar modelos utilizando Frameworks e Linguagens para realizar previsões em escala.
      - PySpark, Scala, Java, Python, C++
    - Definir formas de processamento.
      - Streaming, Clusters.
    - Monitoramento e avaliação de resultados.



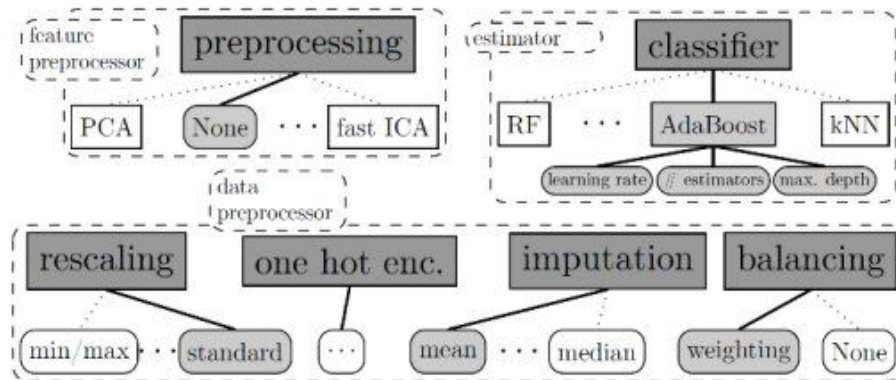
# Introdução a AutoML

AutoML - Machine Learning Automatizado.

Fornece métodos, técnicas e processos para fazer Machine Learning disponível a profissionais não experts, melhorando resultados com menos esforço.

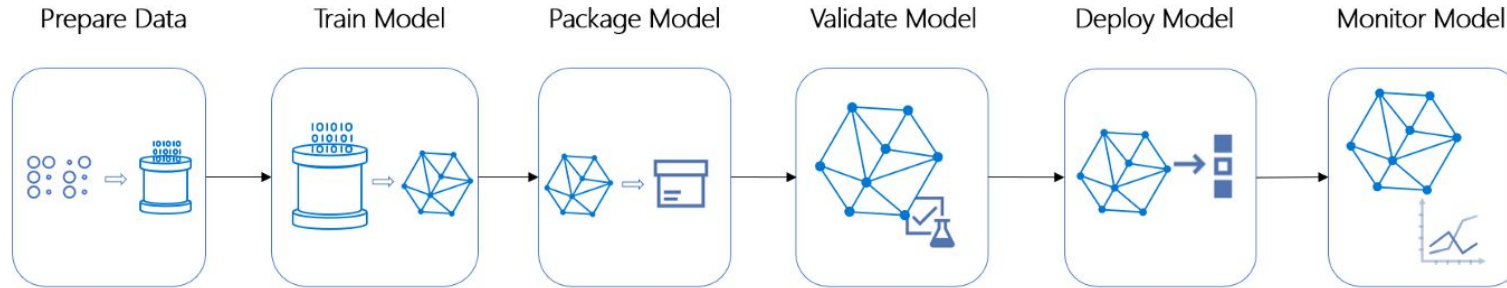
Automação de tarefas como:

- ☐ Pré-processar e limpar dados.
- ☐ Selecionar e construir features apropriadas.
- ☐ Selecionar o melhor modelo.
- ☐ Otimizar hiperparâmetros do modelo.
- ☐ Projetar a topologia de redes neurais.
- ☐ Desenvolver pipelines para predição.
- ☐ Analisar os resultados obtidos.



# Machine Learning Pipeline

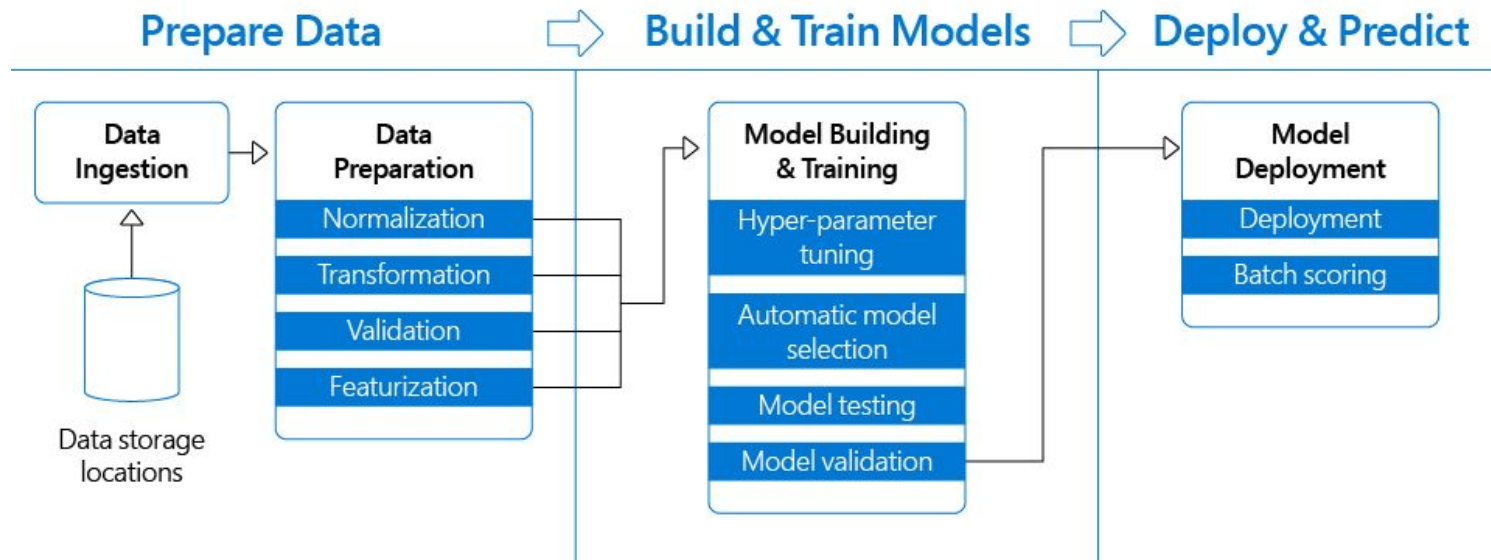
Etapas de um Pipeline para disponibilização de um modelo de Machine Learning.





# Machine Learning Pipeline

Pycaret nos permite automatizar todas as etapas de um Pipeline para disponibilização de um modelo de Machine Learning.





# Machine Learning Pipeline

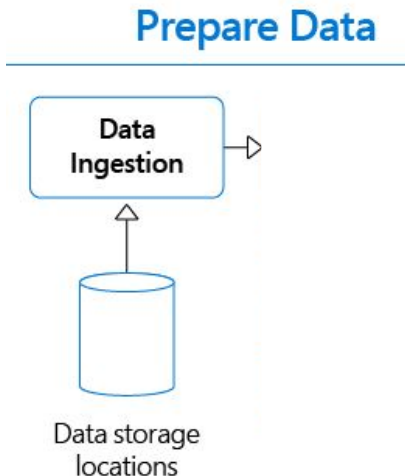
## Preparação de Dados e Ingestão

### Detecção de tipos de colunas.

- Boolean.
- Numérica discreta/contínua.
- Textual.
- Label/Target

### Detecção de tarefas.

- Classificação binária.
- Regressão.
- Clustering.
- Ranking.



# Machine Learning Pipeline

## Feature Engineer

- Seleção de Features.
- Extração de Features.
- Meta learning e transfer learning
- Detecção e manipulação de dados enviesados.
- Manipulação de valores faltantes.

### Prepare Data

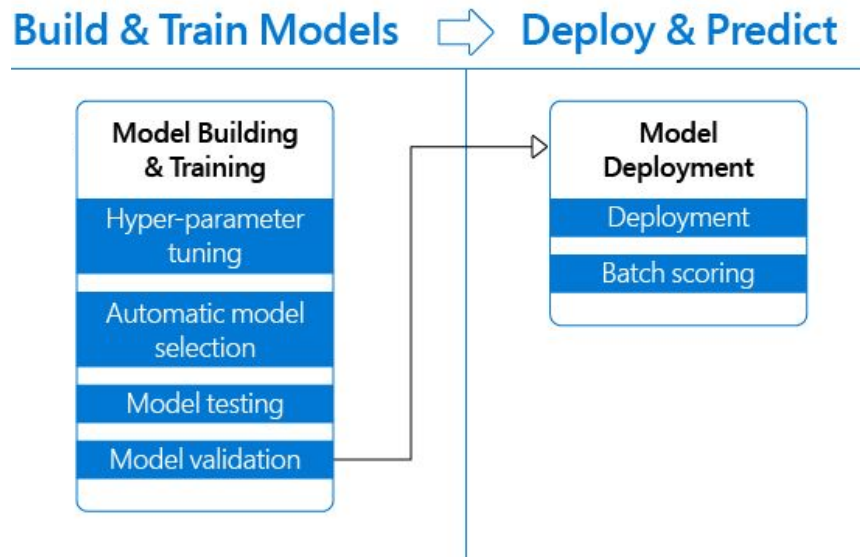
---



# Machine Learning Pipeline

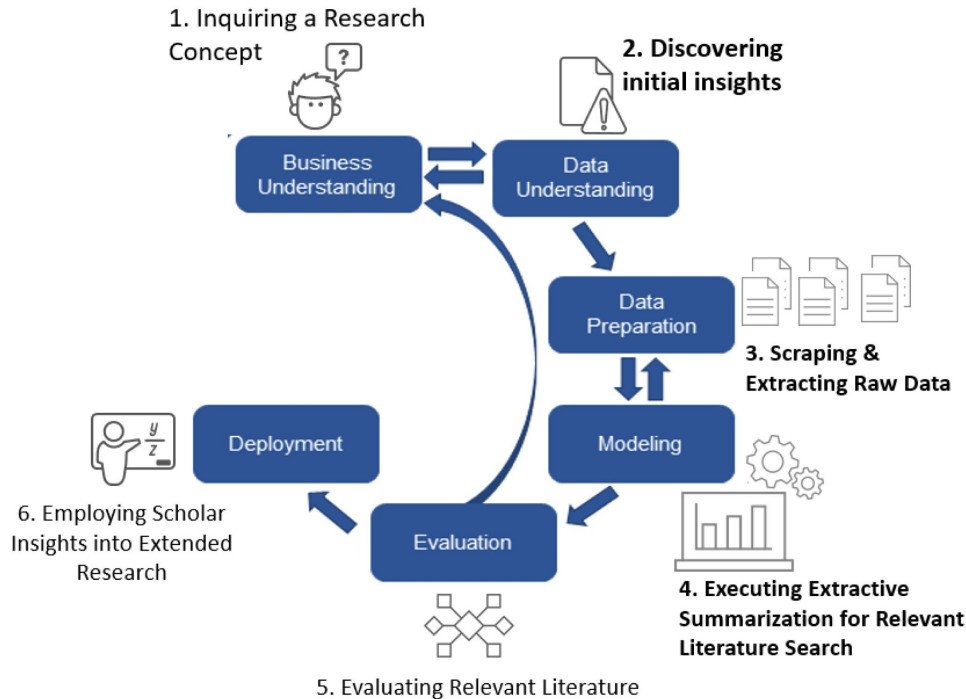
## Seleção de Modelos e Otimização de Hiperparâmetros

- Seleção de modelos.
- Otimização de hiperparâmetros.
- Seleção de métricas e validação.
- Deploy e avaliação.



# O que NÃO é AutoML

AutoML não é Auto Data Science.



# PYCARET



Data  
Preparation



Model  
Training



Hyperparameter  
Tuning



Analysis &  
Interpretability



Model  
Selection



Experiment  
Logging



"PyCaret is an open-source, low-code machine learning library in Python that aims to reduce the cycle time from hypothesis to insights.

It is well suited for seasoned data scientists who want to increase the productivity of their ML experiments by using PyCaret in their workflows or for citizen data scientists and those new to data science with little or no background in coding.

PyCaret allows you to go from preparing your data to deploying your model within seconds using your choice of notebook environment."



# Pycaret

Biblioteca Python Open Source e Low Code para trabalhar com Machine Learning.

Pycaret automatiza tarefas para resolução de problemas utilizando Machine Learning.

- Classificação.
- Regressão.
- Clustering.
- Detecção de Anomalias.
- Processamento de Linguagem Natural.
- Mineração de regras de associação.



Data  
Preparation



Model  
Training



Hyperparameter  
Tuning



Analysis &  
Interpretability



Model  
Selection



Experiment  
Logging



# Pycaret

Pycaret encapsula funções, técnicas, métodos e algoritmos de outras bibliotecas e frameworks como:

- ☐ CatBoost.
- ☐ LightGBM.
- ☐ XG Boost.
- ☐ Scikit-learn.
- ☐ Spacy.
- ☐ Optuna.
- ☐ Ray.

**PYCARET**



CatBoost



LightGBM

**XGBoost**



**spaCy**



OPTUNA



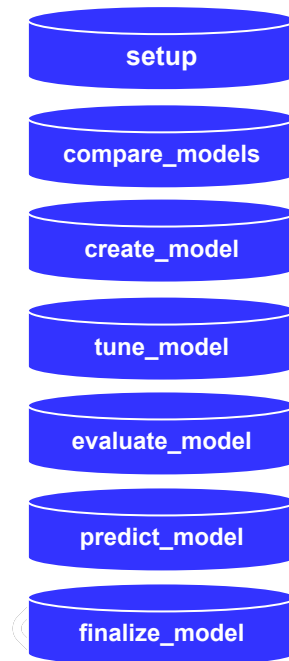
RAY

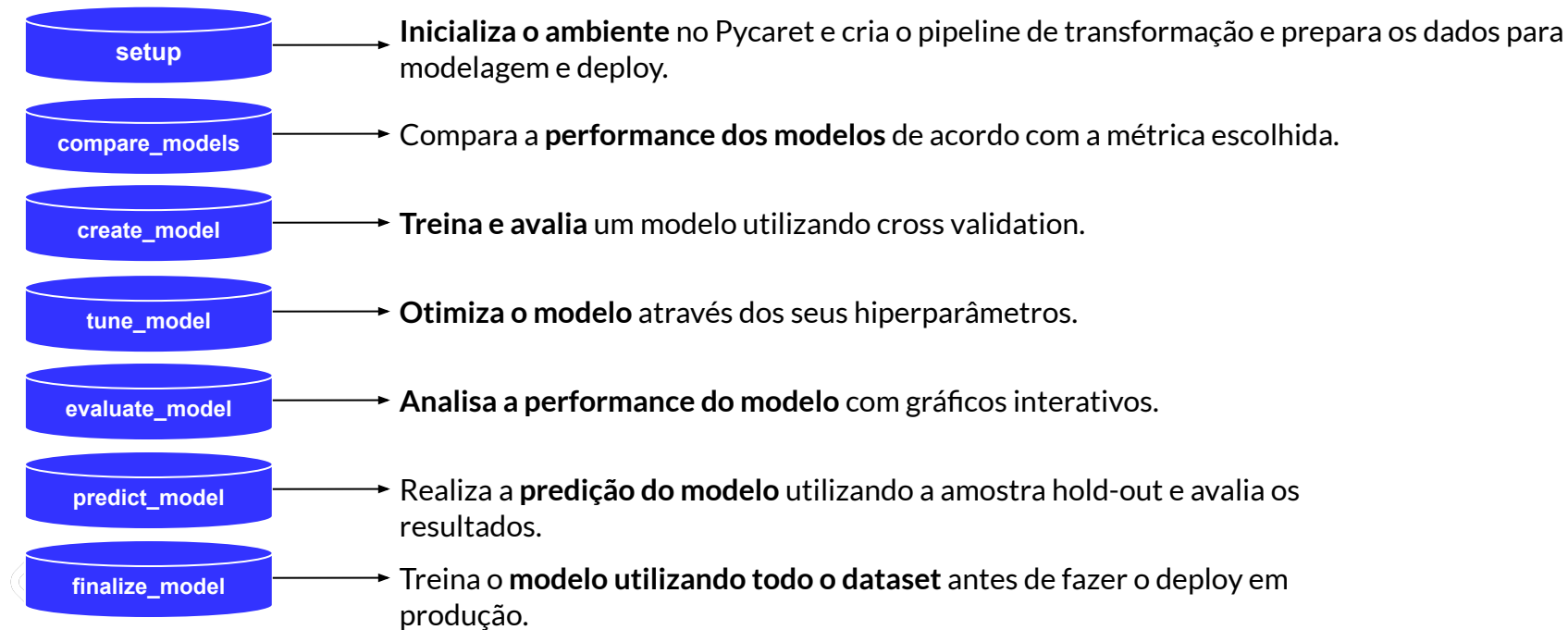


**HANDS ON!**



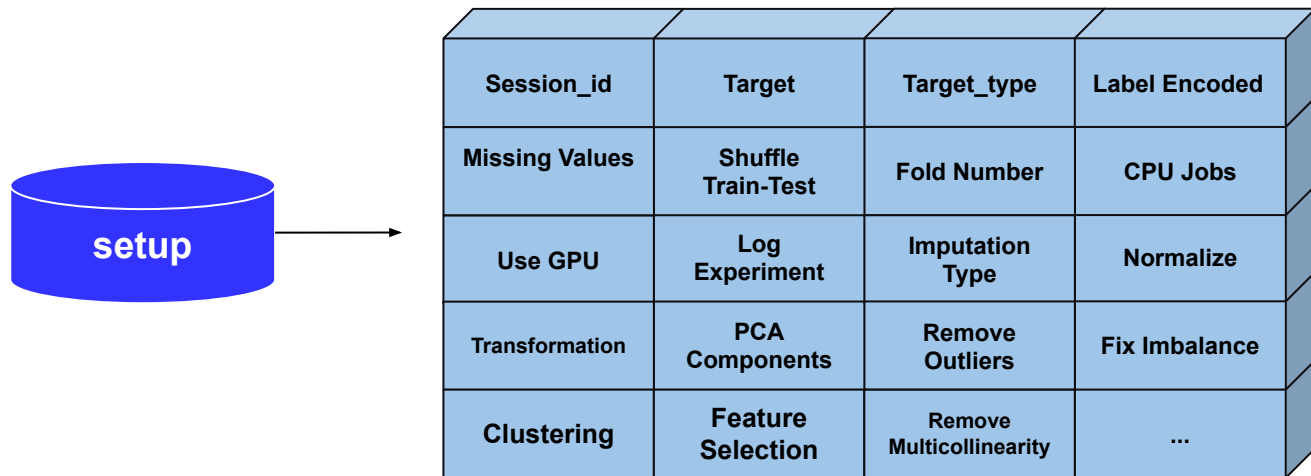
## Funções para implementar o Pipeline de Machine Learning do Pycaret





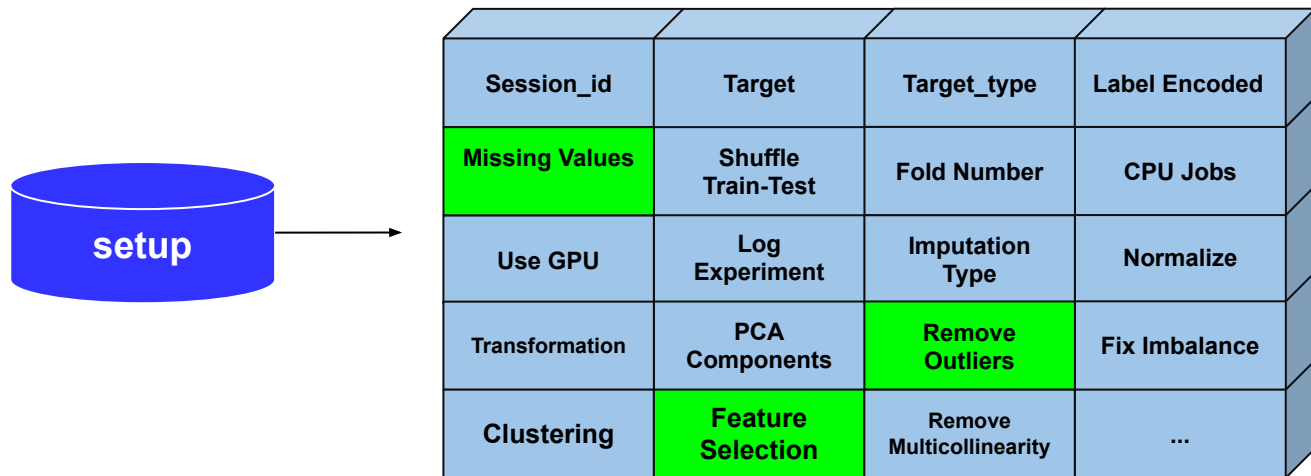


**Inicializa o ambiente** no Pycaret e cria o pipeline de transformação e prepara os dados para modelagem e deploy.



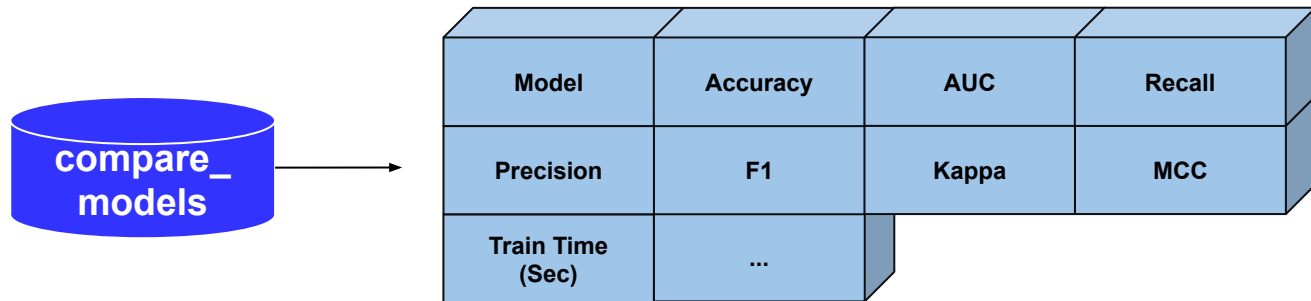
# PYCARET

**Inicializa o ambiente** no Pycaret e cria o pipeline de transformação e prepara os dados para modelagem e deploy.





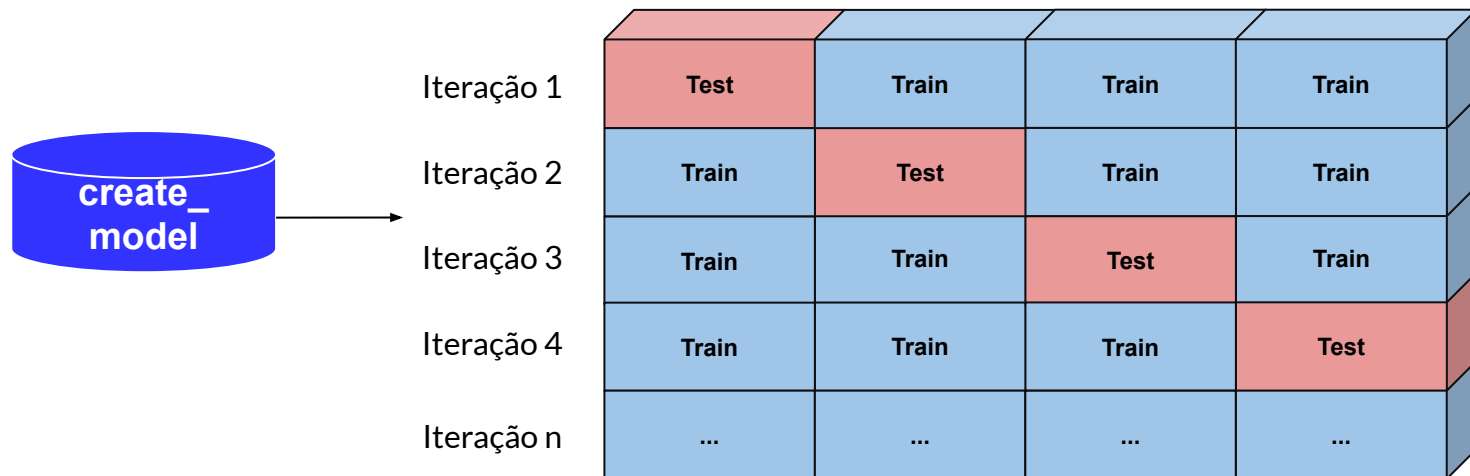
Esta função **treina e compara a performance de todos os modelos** para a determinada tarefa escolhida (Classificação, Regressão, Clustering etc) utilizando cross validation e retorna as métricas para avaliação e o tempo de treinamento.







Esta função **treina o modelo especificado** utilizando cross validation e retorna as métricas para avaliação..



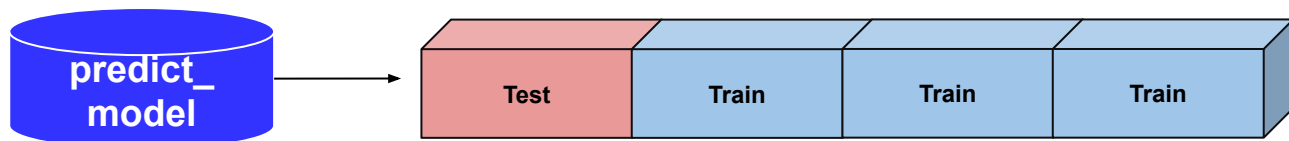
# PYCARET

Esta função **otimiza o modelo através dos seus hiperparâmetros** utilizando diferentes técnicas e ferramentas de otimização como scikit-learn, scikit-optimize, hyperopt e optuna.



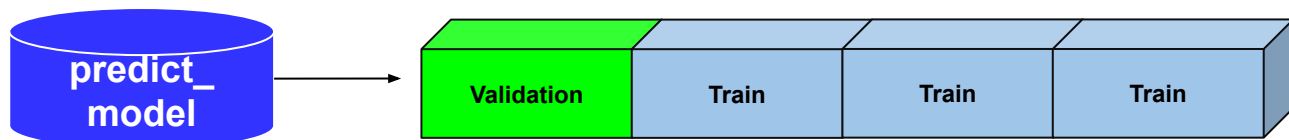


Esta função **realiza a predição** utilizando o modelo especificado utilizando a amostra separada de teste.





Esta função realiza a predição utilizando o modelo especificado utilizando a amostra separada como um conjunto de validação.





Esta função realiza a predição utilizando o modelo especificado utilizando a amostra separada como um conjunto de validação.

