



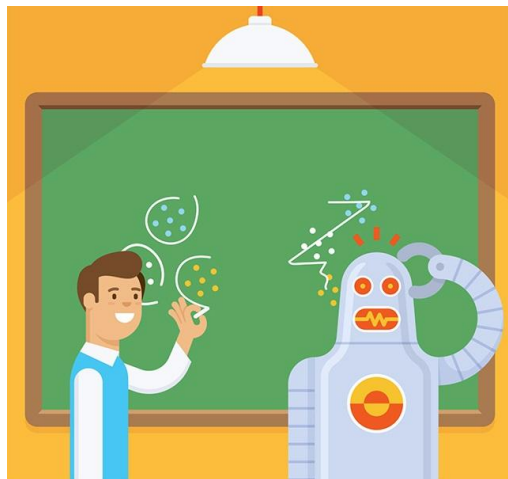
**UFAM**

# QuickAutoML: Uma ferramenta para treinamento automatizado de modelos de aprendizado de máquina

Guilherme Siqueira, Diego Kreutz, Eduardo Feitosa, Gustavo Rodrigues

# Aprendizado de máquina

- Identifica padrões através de **exemplos**
- Caso de uso: detecção de aplicações malignas em Android



# Etapas do aprendizado de máquina

Coleta de  
dados

Preparação  
dos dados

Seleção de  
algoritmos

Treinamento

Ajuste do  
modelo

# Etapas do aprendizado de máquina

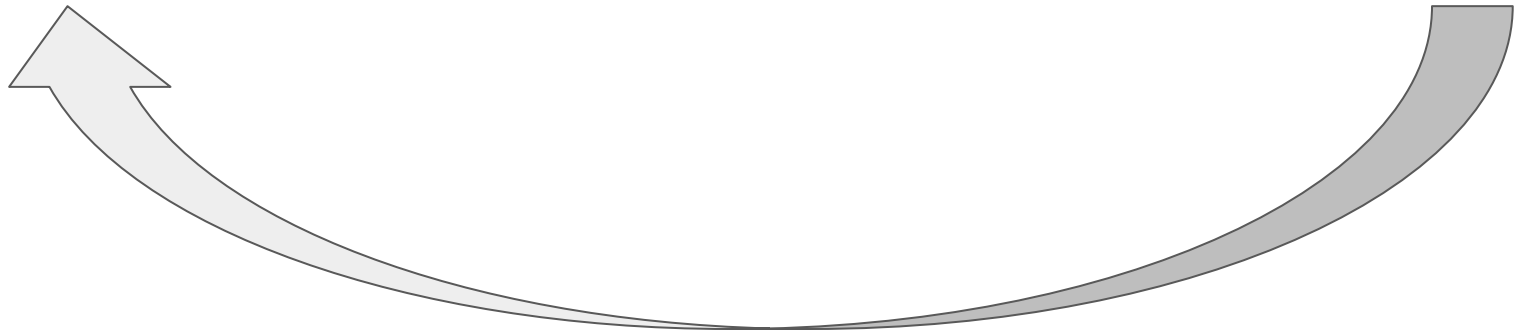
Coleta de  
dados

Preparação  
dos dados

Seleção de  
algoritmos

Treinamento

Ajuste do  
modelo



# Etapas do aprendizado de máquina

## 1. Coleta de dados

- Leitura de bancos de dados
- Aquisição de imagens, vídeos, arquivos Excel, etc.
- Leitura de sensores



# Etapas do aprendizado de máquina

## 2. Preparação dos dados

- Tratamento de valores ausentes
- Conversão de tipos de dado
- Tratamento de erros nos valores

# Etapas do aprendizado de máquina

## 2. Preparação dos dados

- Tratamento de valores ausentes
- Conversão de tipos de dado
- Tratamento de erros nos valores

1 != "1"

```
{  
  "nome": "John",  
  "idade": 300  
}
```

# Etapas do aprendizado de máquina

## 3. Seleção de algoritmos

Florestas Aleatórias?  
K Vizinhos Próximos?  
Máquinas de Vetor de  
Suporte?

...

- Avaliar tamanho do conjunto de dados
- Avaliar interpretabilidade
- Medir a distribuição dos dados



# Etapas do aprendizado de máquina

## 4. Treinamento

- Definir métrica
- Avaliar o enviesamento
- Validar os resultados

# Etapas do aprendizado de máquina

## 4. Treinamento

Acurácia?  
Revocação?  
Precisão?



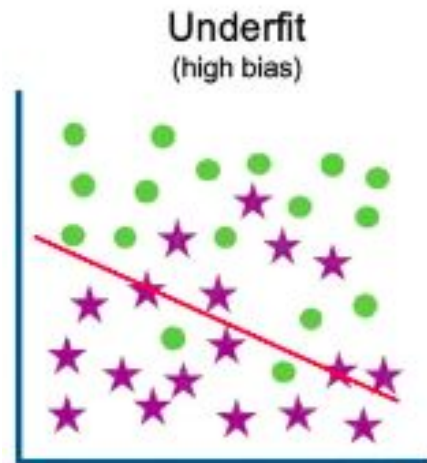
...

- Definir **métrica**
- Avaliar o enviesamento
- Validar os resultados

# Etapas do aprendizado de máquina

## 4. Treinamento

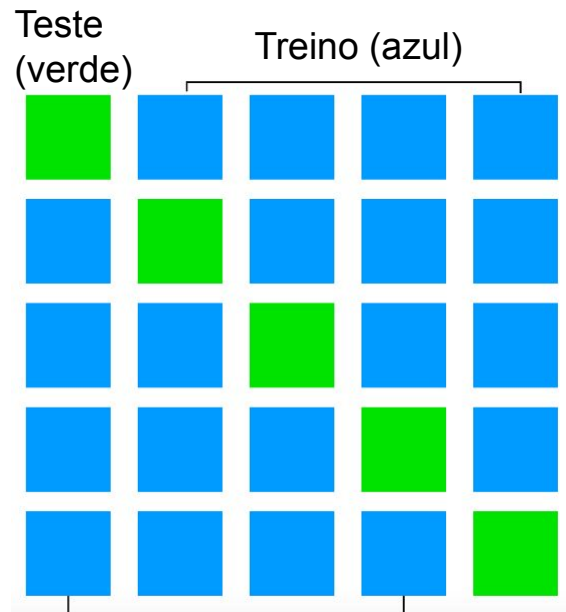
- Definir métrica
- Avaliar o **enviesamento**
- Validar os resultados



# Etapas do aprendizado de máquina

## 4. Treinamento

- Definir métrica
- Avaliar o enviesamento
- Validar os **resultados**



# Etapas do aprendizado de máquina

## 5. Refinamento do modelo

- Ajuste de hiper-parâmetros

# Etapas do aprendizado de máquina

## 5. Refinamento do modelo

- Ajuste de hiper-parâmetros

RandomForest

n\_estimators = ?  
max\_depth = ?  
min\_samples\_split = ?  
max\_features = ?  
max\_leaf\_nodes = ?  
criterion = ?

...

# Motivação

- Complexidade do processo de treinamento
- Necessidade de soluções para automatização do treinamento (AutoML)

# Proposta de ferramenta

Coleta de  
dados

Preparação  
dos dados

Seleção de  
algoritmos

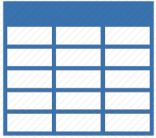
Treinamento

Refinamento  
do modelo

QuickAutoML:  
Simplificar e automatizar

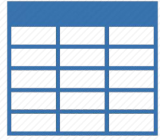


# Implementação

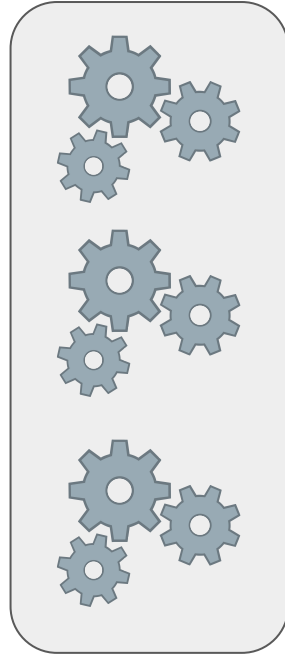
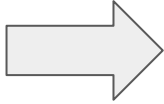


Conjunto  
de dados

# Implementação

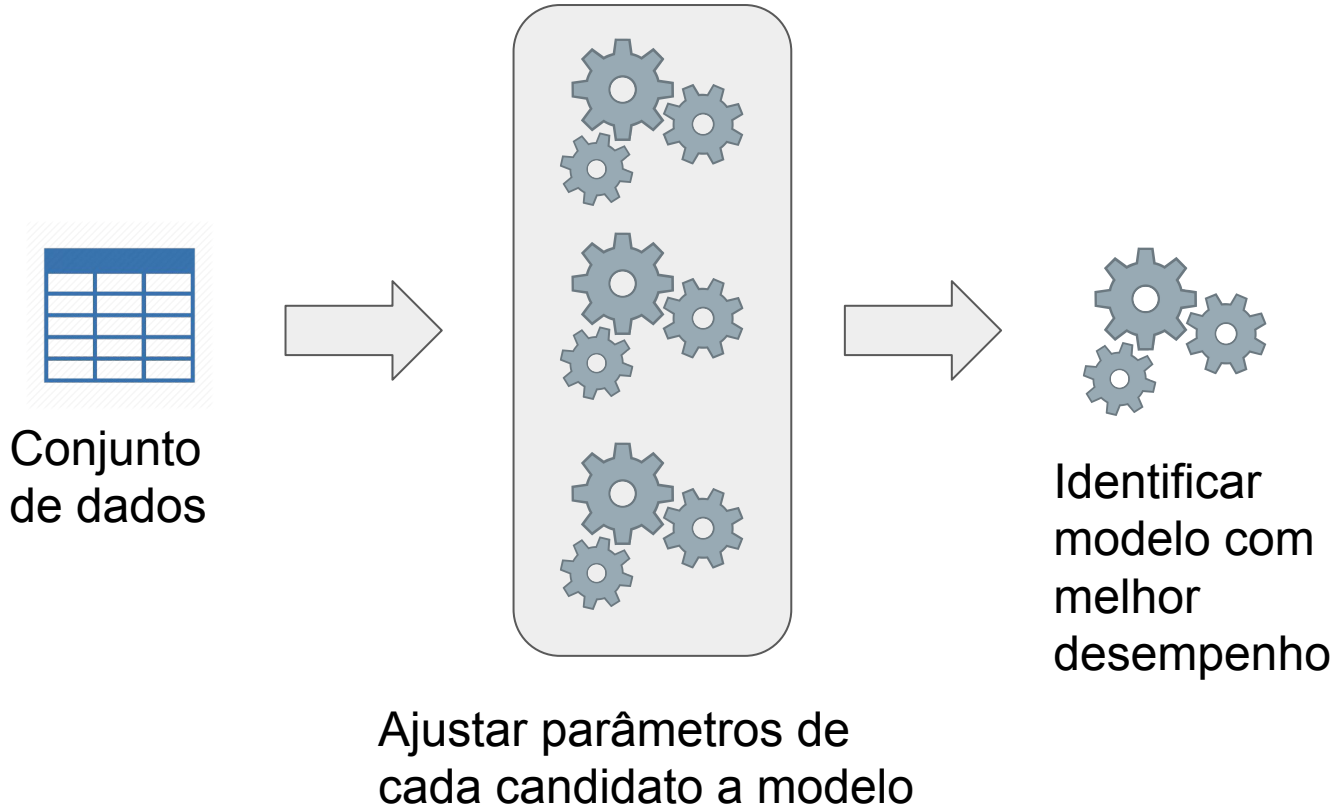


Conjunto  
de dados

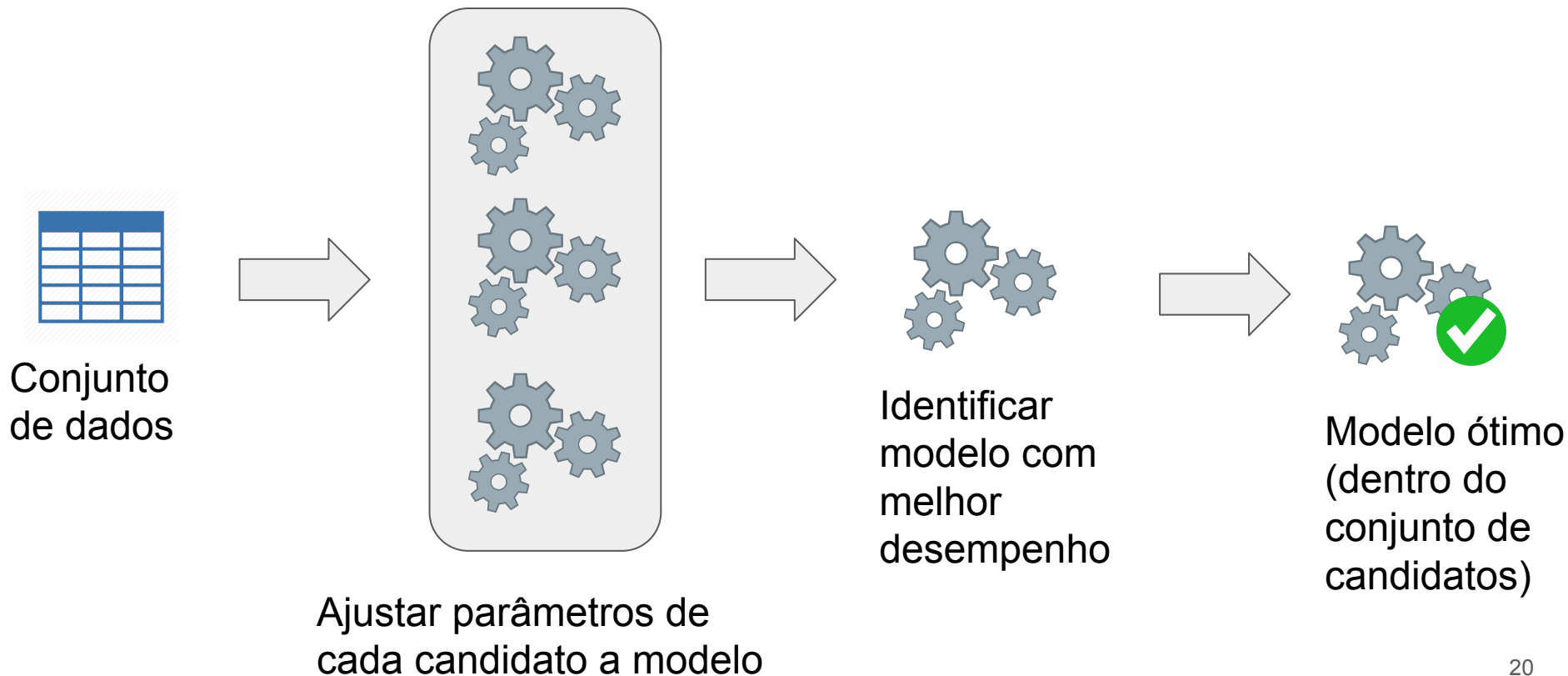


Ajustar parâmetros de  
cada candidato a modelo

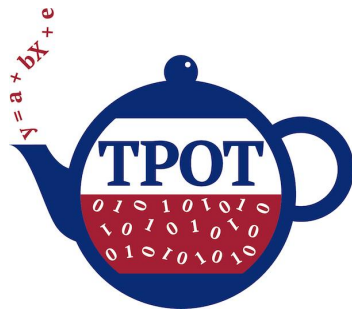
# Implementação



# Implementação



**TPOT:** Uma ferramenta de otimização de pipelines de aprendizado de máquina



# Avaliação

Nome	Número de amostras	Número de características
new_thyroid	215	5
hypothyroid	3163	25
dermatology	366	34
hayes_roth	160	4
dis	3772	29
penguins	344	7

Tabela 1: Datasets utilizados para avaliação

## Resultados obtidos

<b>dataset</b>	<b>Tempo (em segundos) - QuickAutoML</b>	<b>Acurácia - QuickAutoML</b>	<b>Tempo (em segundos) - TPOT</b>	<b>Acurácia - TPOT</b>
new_thyroid	3.6433	0.9535	29.8389	0.9535
hypothyroid	7.1176	0.9858	158.2114	0.9905
dermatology	4.2010	0.9730	54.8852	0.9595
hayes_roth	3.9715	0.6875	29.4409	0.75
dis	7.6694	0.9881	209.2988	0.9907
penguins	3.3933	0.9551	40.5660	0.9701

**Tabela 2: Resultados obtidos na comparação entre QuickAutoML e TPOT**

# Resultados obtidos

Resultados  
semelhantes

	<b>Tempo (em segundos) - QuickAutoML</b>	<b>Acurácia - QuickAutoML</b>	<b>Tempo (em segundos) - TPOT</b>	<b>Acurácia - TPOT</b>
new_thyroid	3.6433	0.9535	29.8389	0.9535
hypothyroid	7.1176	0.9858	158.2114	0.9905
dermatology	4.2010	0.9730	54.8852	0.9595
hayes_roth	3.9715	0.6875	29.4409	0.75
dis	7.6694	0.9881	209.2988	0.9907
penguins	3.3933	0.9551	40.5660	0.9701

**Tabela 2: Resultados obtidos na comparação entre QuickAutoML e TPOT**



# Resultados obtidos

Grandes  
diferenças de  
tempo

	<b>Tempo (em segundos) - QuickAutoML</b>	<b>Acurácia - QuickAutoML</b>	<b>Tempo (em segundos) - TPOT</b>	<b>Acurácia - TPOT</b>
new_thyroid	3.6433	0.9535	29.8389	0.9535
hypothyroid	7.1176	0.9858	158.2114	0.9905
dermatology	4.2010	0.9730	54.8852	0.9595
hayes_roth	3.9715	0.6875	29.4409	0.75
dis	7.6694	0.9881	209.2988	0.9907
penguins	3.3933	0.9551	40.5660	0.9701

**Tabela 2: Resultados obtidos na comparação entre QuickAutoML e TPOT**

# Resultados obtidos

Resultado  
melhor em  
menos tempo

	<b>Tempo (em segundos) - QuickAutoML</b>	<b>Acurácia - QuickAutoML</b>	<b>Tempo (em segundos) - TPOT</b>	<b>Acurácia - TPOT</b>
new_thyroid	3.6433	0.9535	29.8389	0.9535
hypothyroid	7.1176	0.9858	158.2114	0.9905
dermatology	4.2010	0.9730	54.8852	0.9595
hayes_roth	3.9715	0.6875	29.4409	0.75
dis	7.6694	0.9881	209.2988	0.9907
penguins	3.3933	0.9551	40.5660	0.9701

**Tabela 2: Resultados obtidos na comparação entre QuickAutoML e TPOT**

## Considerações Finais

- Balanço entre desempenho e tempo de execução

## Melhorias futuras

- Suporte a outros algoritmos
- Implementação de etapas de processamento de dados
- Utilização de interface gráfica de usuário

# Obrigado!!



**UFAM**

