

COMO SE TORNAR UM MESTRE POKÉMON UTILIZANDO **MACHINE LEARNING**

CINTHIA PESSANHA



CINTHIA PESSANHA

- +10 anos de experiência em desenvolvimento de software na plataforma .NET para a indústria oil & gas.
- Graduada em Ciência da Computação
Universidade Candido Mendes – Campos/RJ
- Mestre em Inteligência Computacional
Universidade Cândido Mendes – Campos/RJ
- Em processo de adoção de um jabuti e tentando fazer meu marido aceitá-lo 😊



cinthiabpessanha

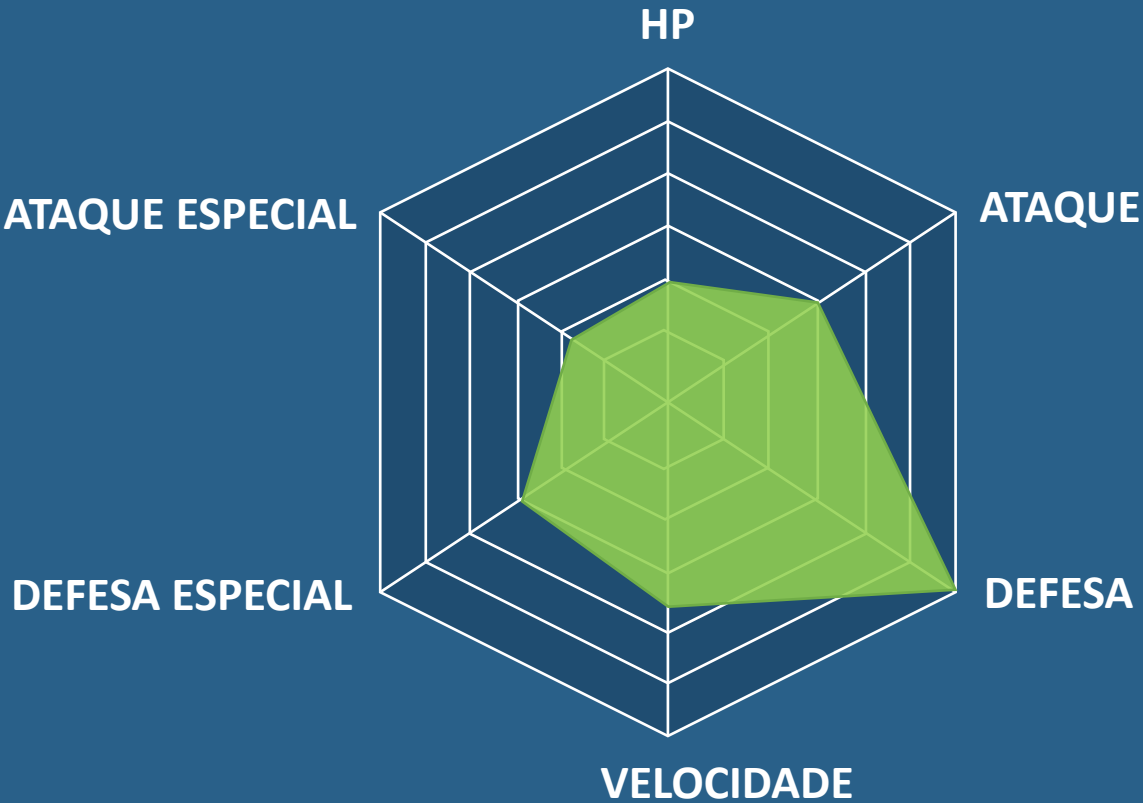
AGENDA

- O que é um Pokémon?
- O problema
- IA – Estado da arte
- Árvores de Decisão
- Random Forest
- Etapas
- Show me THE CODE
- Referências



POKÉMON

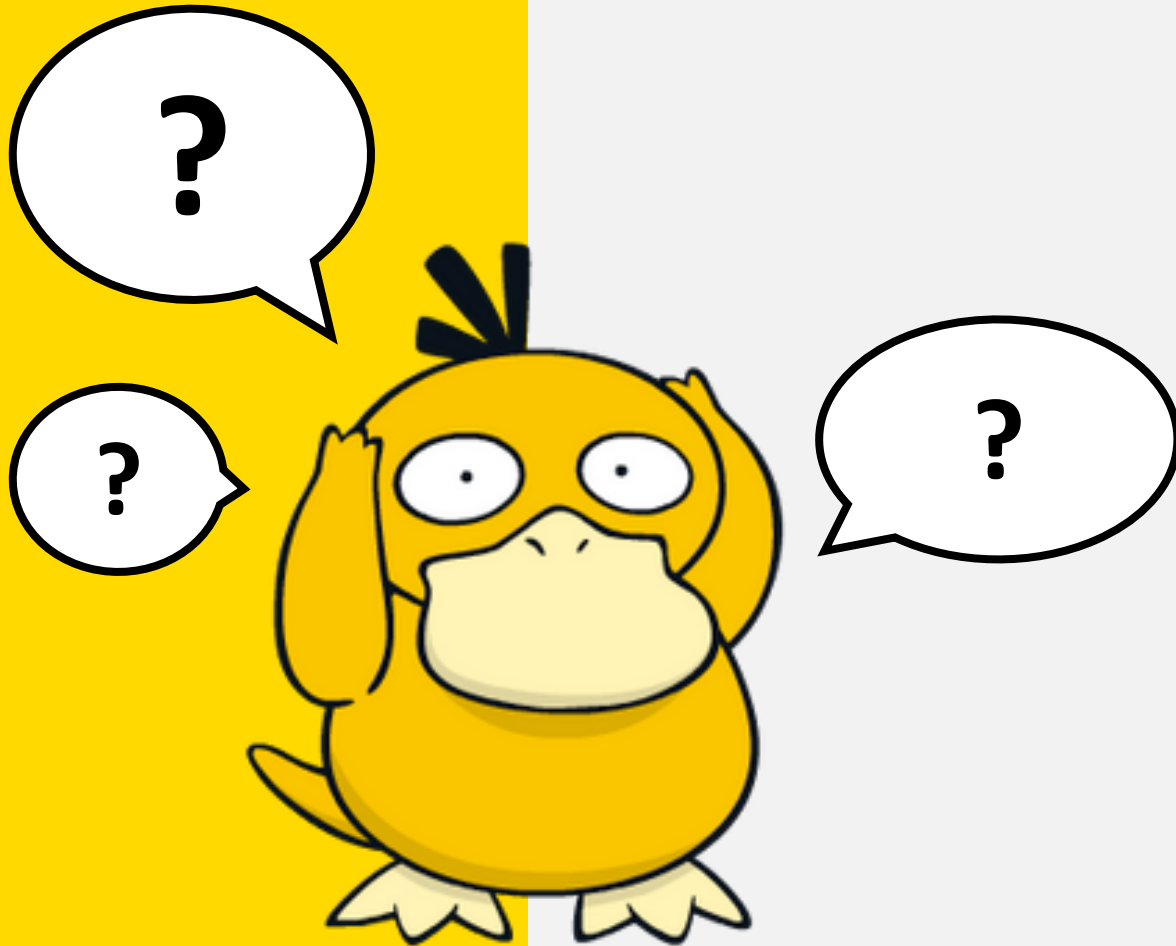
CARACTERÍSTICAS DE UM POKÉMON



TIPO	VANTAGEM	DESVANTAGEM
Aço		
Água		
Dragão		
Elétrico		
Fada		
Fantasama		
Fogo		
Gelo		
Inseto		
Lutador		
Normal		
Pedra		
Planta		
Psíquico		
Sombrio		
Terra		
Venenoso		
Voador		

O PROBLEMA

**DADOS 2 POKÉMONS, QUAL TEM MAIS
CHANCE DE GANHAR UMA BATALHA?**





O QUE ACHAM DE UTILIZARMOS
INTELIGÊNCIA ARTIFICIAL?

ESTADO DA ARTE

DATA
SCIENCE

DEEP
LEARNING

REDES
NEURAIS



INTELIGÊNCIA
ARTIFICIAL

MACHINE
LEARNING

Artificial Intelligence (AI)

Reproduction of human intellectual activities



Component technologies

Expert systems

Robotics

Human
interfaces

⋮

Machine learning



Algorithms

SVM

Random
Forest

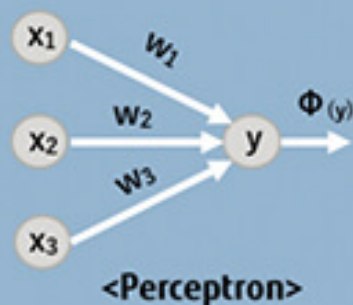
K-means

⋮

Learning of rules and patterns hidden in data

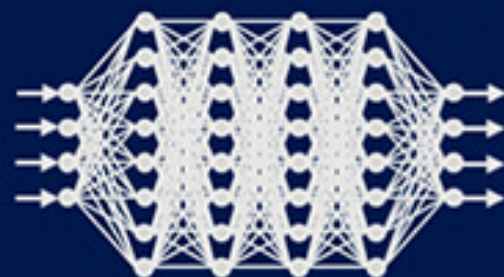
Neural networks

Learning model that emulates brain activities



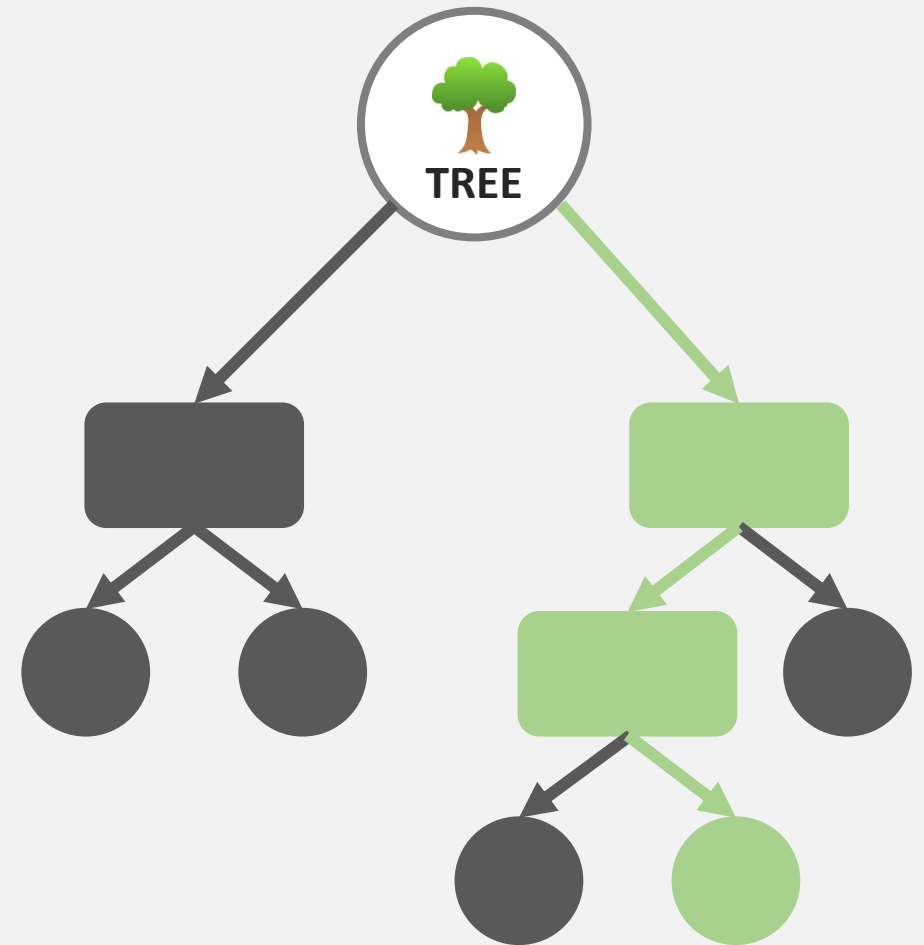
Modeling of neural
signaling

Deep Learning Learning using multi-layer neural networks



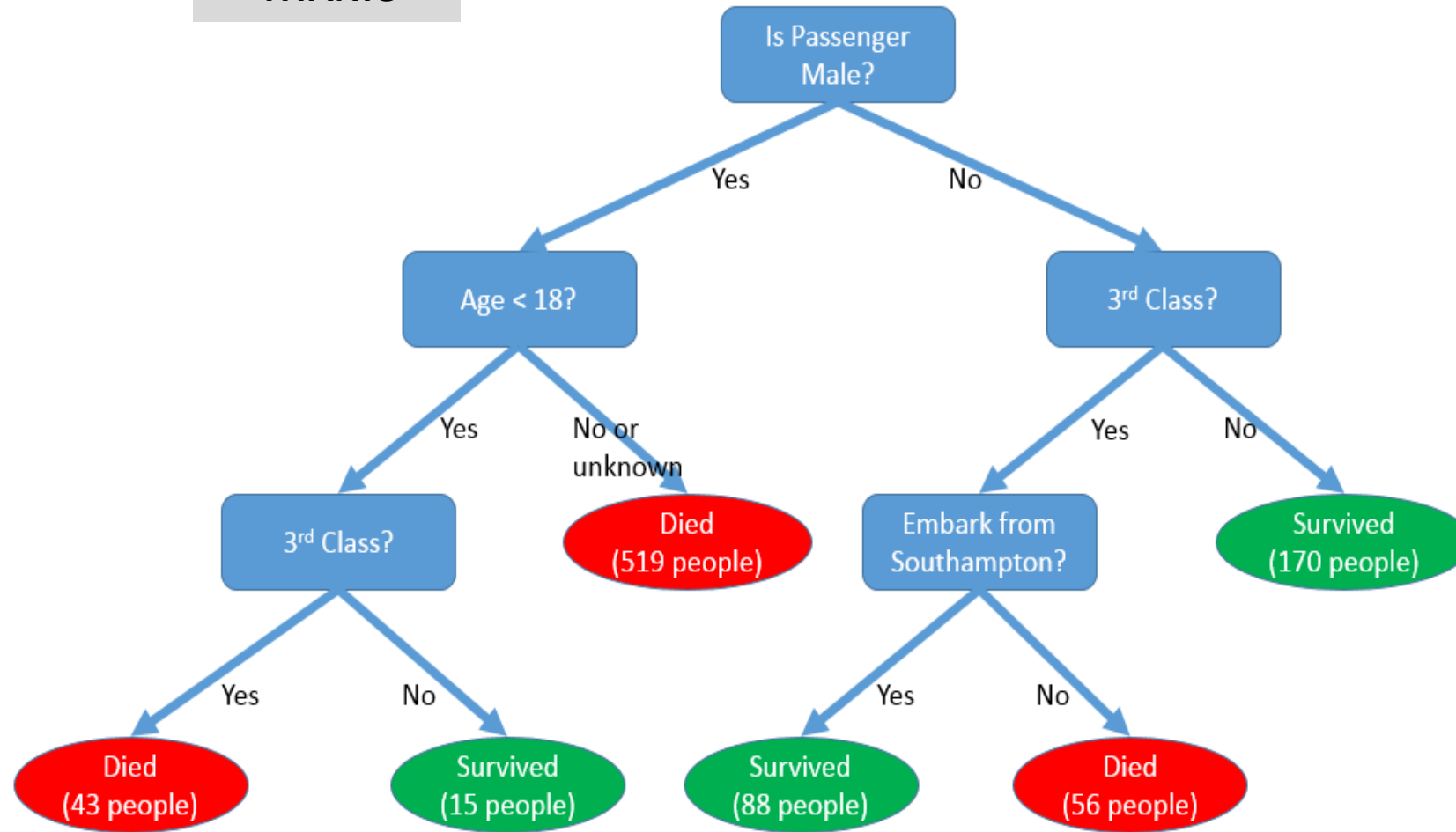
ÁRVORE DE DECISÃO

- Ferramenta de suporte a decisão
- Extrai regras a partir de dados existentes
- Problema: Risco de **Overfitting**

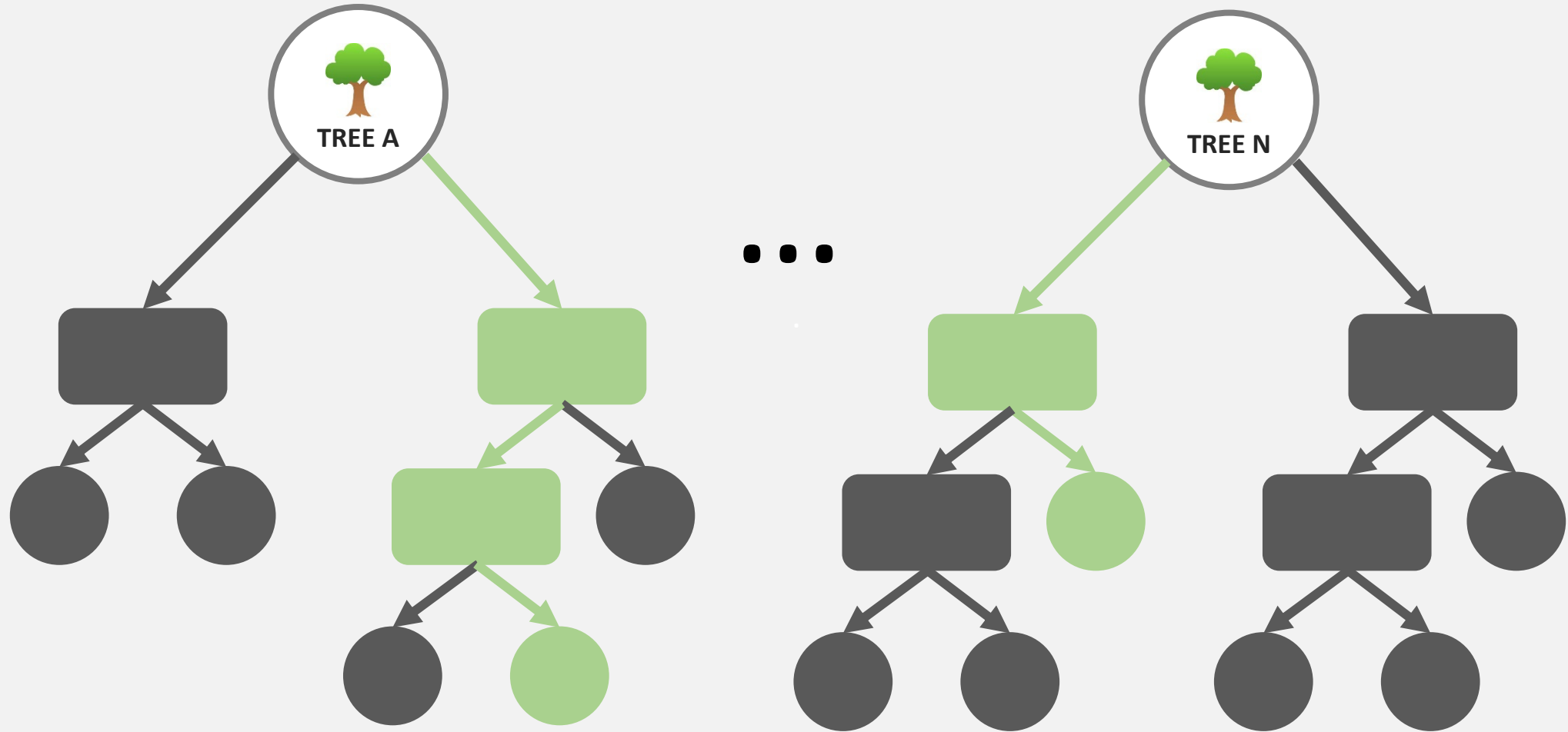


ÁRVORE DE DECISÃO

TITANIC



RANDOM FOREST



RANDOM FOREST

- Várias árvores de decisão são geradas
- Redução significativa de risco de overfitting
- Geração das árvores:
 - Seleção aleatória de algumas features
 - Seleção da feature mais adequada para nó raiz
 - Geração dos nós filhos
 - Repete os passos acima até que se atinja a quantidade de árvores desejada

RANDOM FOREST

- Previsão
 - Dado um conjunto de entrada, cada árvore, com seu conjunto de features selecionadas aleatoriamente, “toma uma decisão”
 - A resposta da Random Forest será a decisão mais “votada”
 - Democracia entre árvores 😊

ETAPAS

**Seleção
dos dados**

**Pré-
processa-
mento**

**Escolha
do
algoritmo**

**Treiname-
nto**

Teste

**Aferição
dos
resultados**

ETAPAS

**Seleção
dos dados**

**Pré-
processa-
mento**

**Escolha
do
algoritmo**

**Treiname-
nto**

Teste

**Aferição
dos
resultados**

- Base de dados não pode ser pequena
 - Lembre-se que parte desses dados será destinado a etapa de testes

ETAPAS

**Seleção
dos dados**

**Pré-
processa
mento**

**Escolha
do
algoritmo**

**Treiname
nto**

Teste

**Aferição
dos
resultados**

- Preenchimento de dados em branco
- Verificar com o especialista como ajustar dados que estão faltando
- Dummies / Encoders para manter ordem de grandeza entre os diferentes tipos de features
- Em caso de imagens, geralmente é necessário submetê-las a algum tratamento
 - Ex: impressões digitais - binarização

ETAPAS

**Seleção
dos dados**

**Pré-
processa-
mento**

**Escolha
do
algoritmo**

**Treiname-
nto**

Teste

**Aferição
dos
resultados**

- A escolha dependerá do tipo de problema a ser resolvido
- Teste vários algoritmos
- Para atingir o modelo que responda com maior acurácia, testar diferentes abordagens com diferentes parâmetros é importante
- A escolha do algoritmo pode exigir ajustes nos dados

ETAPAS

**Seleção
dos dados**

**Pré-
processa
mento**

**Escolha
do
algoritmo**

**Treiname
nto**

Teste

**Aferição
dos
resultados**

- Geração do modelo

ETAPAS

**Seleção
dos dados**

**Pré-
processa-
mento**

**Escolha
do
algoritmo**

**Treiname-
nto**

Teste

**Aferição
dos
resultados**

- O percentual de dados de teste é variável
- Teste diversos percentuais

Treinamento

Teste

ETAPAS

**Seleção
dos dados**

**Pré-
processa-
mento**

**Escolha
do
algoritmo**

**Treiname-
nto**

Teste

**Aferição
dos
resultados**

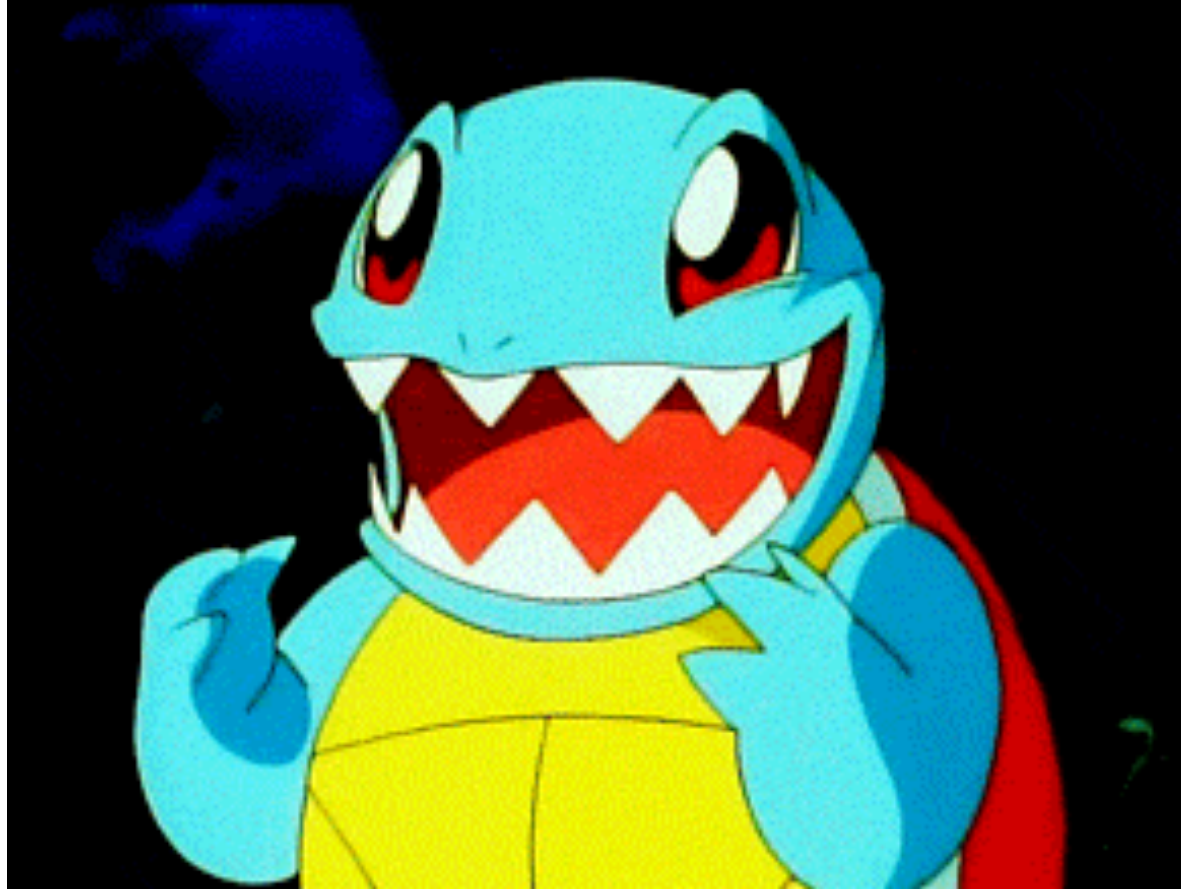
ETAPAS



FERRAMENTAS



SHOW ME THE CODE



REFERÊNCIAS

- Fonte de dados
 - Kaggle
 - O governo brasileiro disponibiliza dados do ENEM e da Polícia Federal
- Linguagem
 - Python
 - Angular + JS (Tensorflow)
 - Pytorch
- Softwares
 - Anaconda
 - Colab
- Cursos
 - <https://developers.google.com/machine-learning/crash-course/ml-intro>
 - <https://pt.coursera.org/learn/machine-learning>
 - School of AI

OBRIGADA!!!

