Introdução ao Python

5. Aprendizado de Máquina

Rodrigo Barbosa de Santis

Sumário

Contextualização

- Nos primórdios do estudo da inteligência artificial, pesquisadores trabalhavam construindo enormes tabelas, enumerando todas os possíveis cenários das condições observadas, e atribuindo ações (saídas) para o sistema.

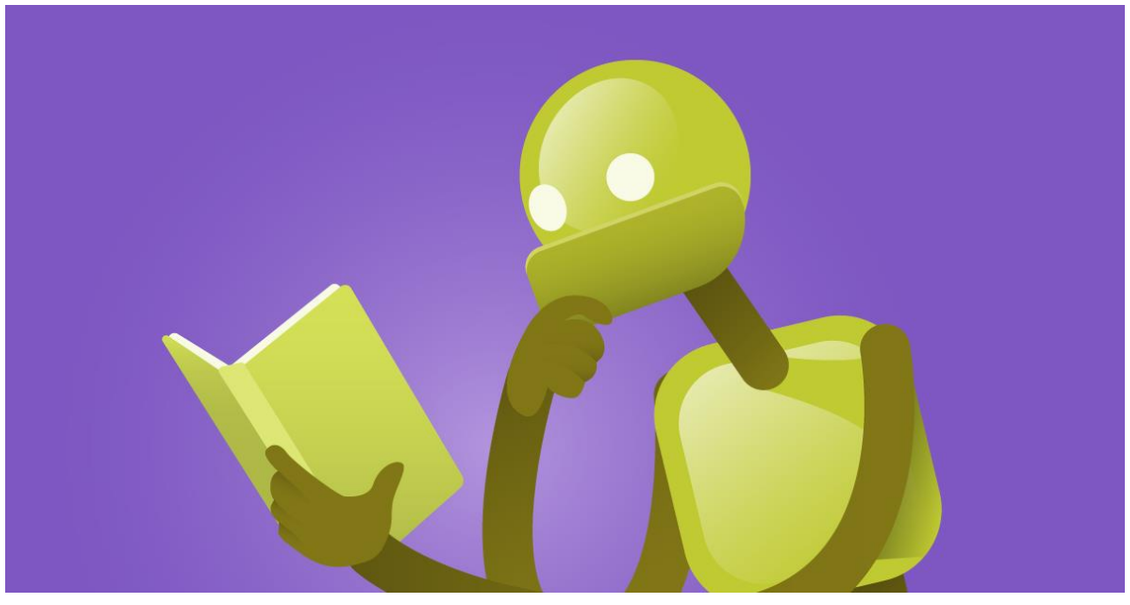
IF				THEN	
Y_1	X_2	Y_2	Y_3	DoS (initial: final)	Incident_ Status
low	low	low	low	(0.50: 0.97)	False
low	low	low	low	(0.50: 0.98)	True
low	low	high	low	(0.50: 0.49)	False
low	low	high	low	(0.50: 0.50)	True
low	low	high	high	(0.50: 0.45)	False
low	low	high	high	(0.50: 0.50)	True

- Este tipo de abordagem ainda existe, porém evoluiu para o que conhecemos como **sistemas de inferência *fuzzy***.

Introdução

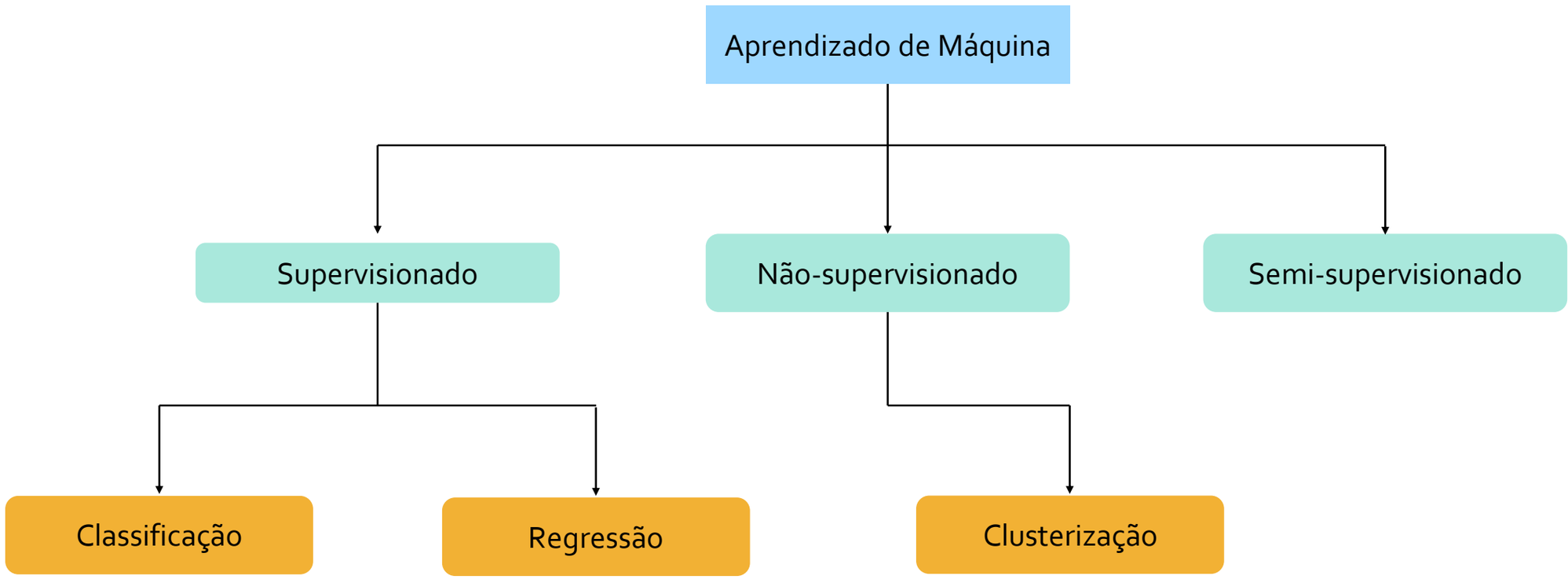
- Já o **aprendizado de máquina** estuda de **algoritmos** e **modelos estatísticos** usados para executar tarefas específicas sem usar instruções explícitas, baseando-se em **padrões** e **inferência**.

- Os modelos de aprendizado de máquina não precisam ser **explicitamente** programados para realizar a tarefa de **classificação** ou **regressão**, como os sistemas de inferência tradicionais precisavam.



Introdução

- Atualmente, a área de aprendizado de máquina é dividida da seguinte forma:



Introdução

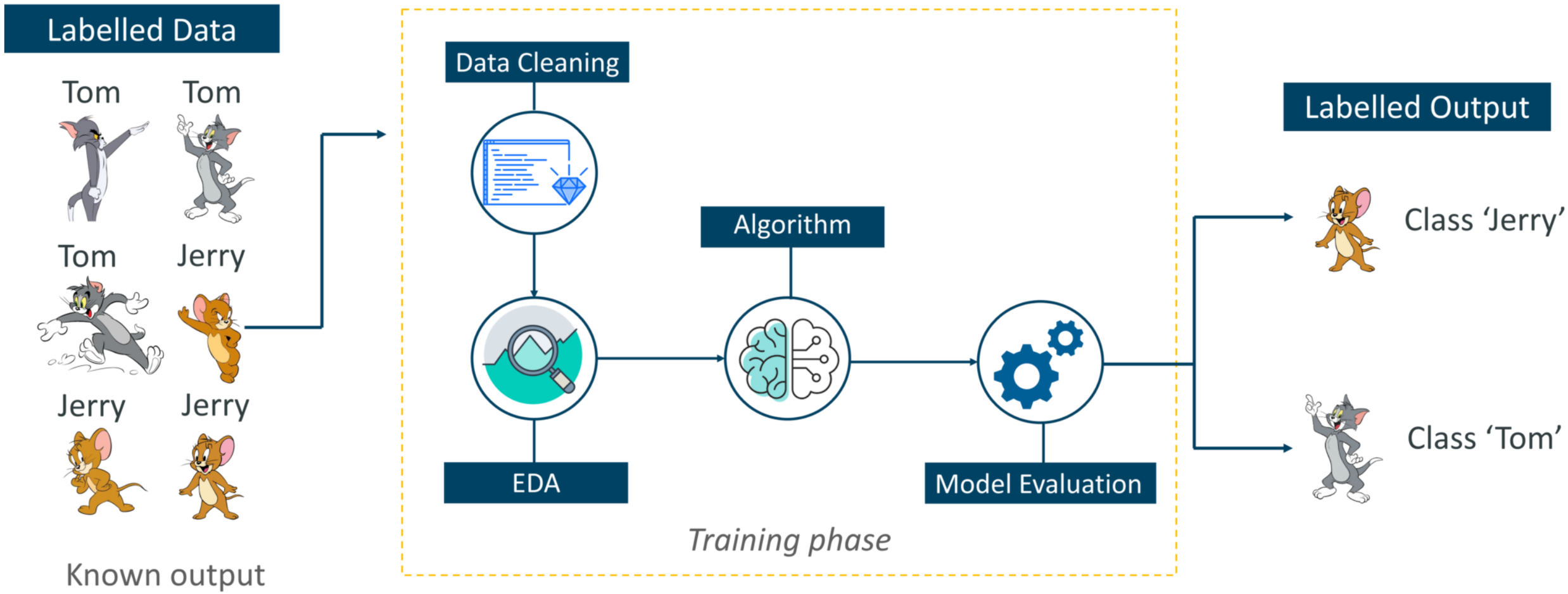
- Atualmente, a área de aprendizado de máquina é dividida da seguinte forma:

  - Aprendizado **supervisionado**: usado para prever uma variável do nosso interesse – classificação para variáveis discretas; regressão para variáveis contínuas. Esta variável precisa estar rotulada, o que em alguns casos é um trabalho oneroso: ex. banco de dados de imagem.

  - Aprendizado **não-supervisionado**: não temos uma variável que queremos prever, estamos apenas explorando e agrupando nossos dados a partir de métricas de distância.

  - Aprendizado **semi-supervisionado**: temos rótulo apenas para parte de nossos dados, usamos clusterização para rotular os dados restantes. Utilizado para facilitar o processo de colocação de rótulos em conjuntos de dados muito grandes.

Aprendizado Supervisionado

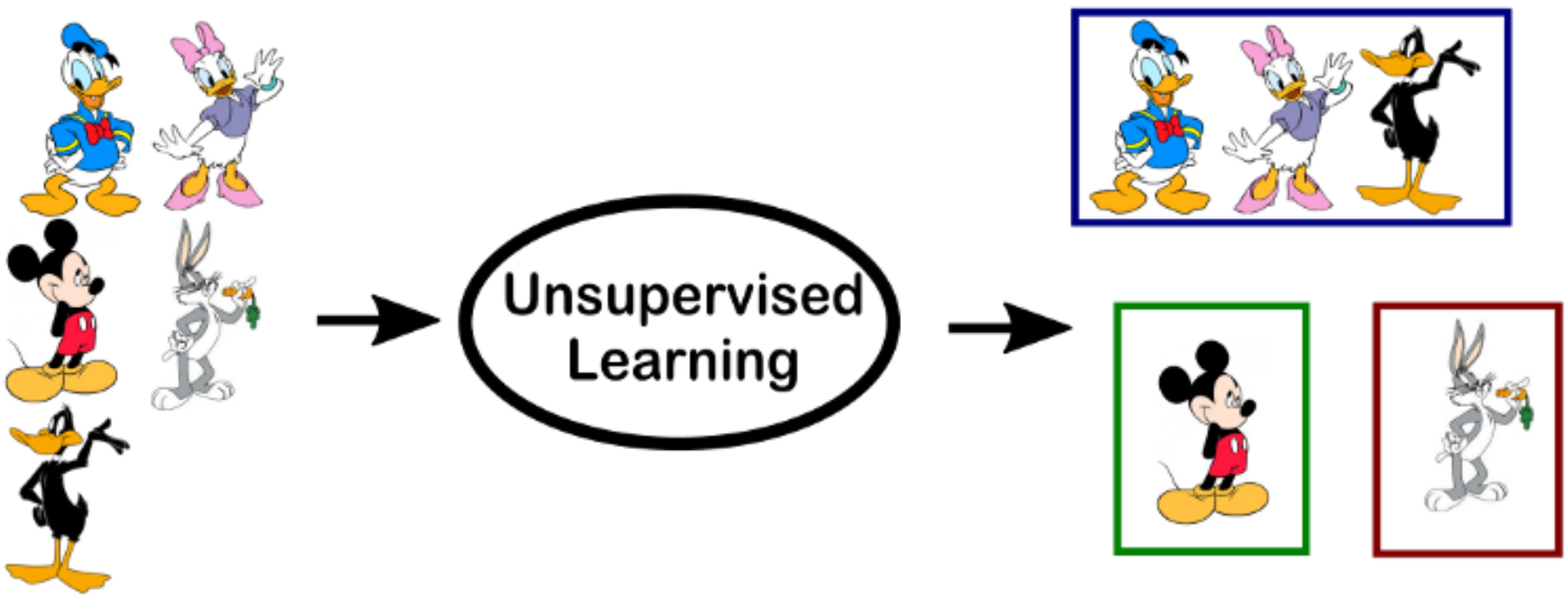


Top 10 Algoritmos

- Regressão Linear
- Regressão Logística
- K-Vizinhos mais Próximos (KNN)
- Support Vector Machine (SVM)
- Decision Tree (CART ou C4.5)
- Comitês
  - Gradient Boosting (GBoost / XGBoost / LightGBM) e Random Forest
- Redes Neurais
  - Multi-layer Perceptron (MLP)
  - Extreme Learning Machine (ELM)
  - Convolutional Neural Networks (CNN) e Recurrent Neural Networks (RNN)

Aprendizado Não-Supervisionado

- Também conhecido como clusterização ou agrupamento, não existem rótulos ou variável a ser prevista. O modelo busca manter as amostras com menor distâncias em N* grupos pré-definidos.



- Na maioria dos métodos, N é um valor a ser definido pelo usuário.

Semi-Supervisionado