

MODEL LIFECYCLE



CONCEITOS BÁSICOS DE MACHINE LEARNING I

DIEGO RODRIGUES DSC

INFNET

PYTHON PARA MACHINE LEARNING

APRESENTAÇÃO DO PROFESSOR

PARTE 1 : TEORIA

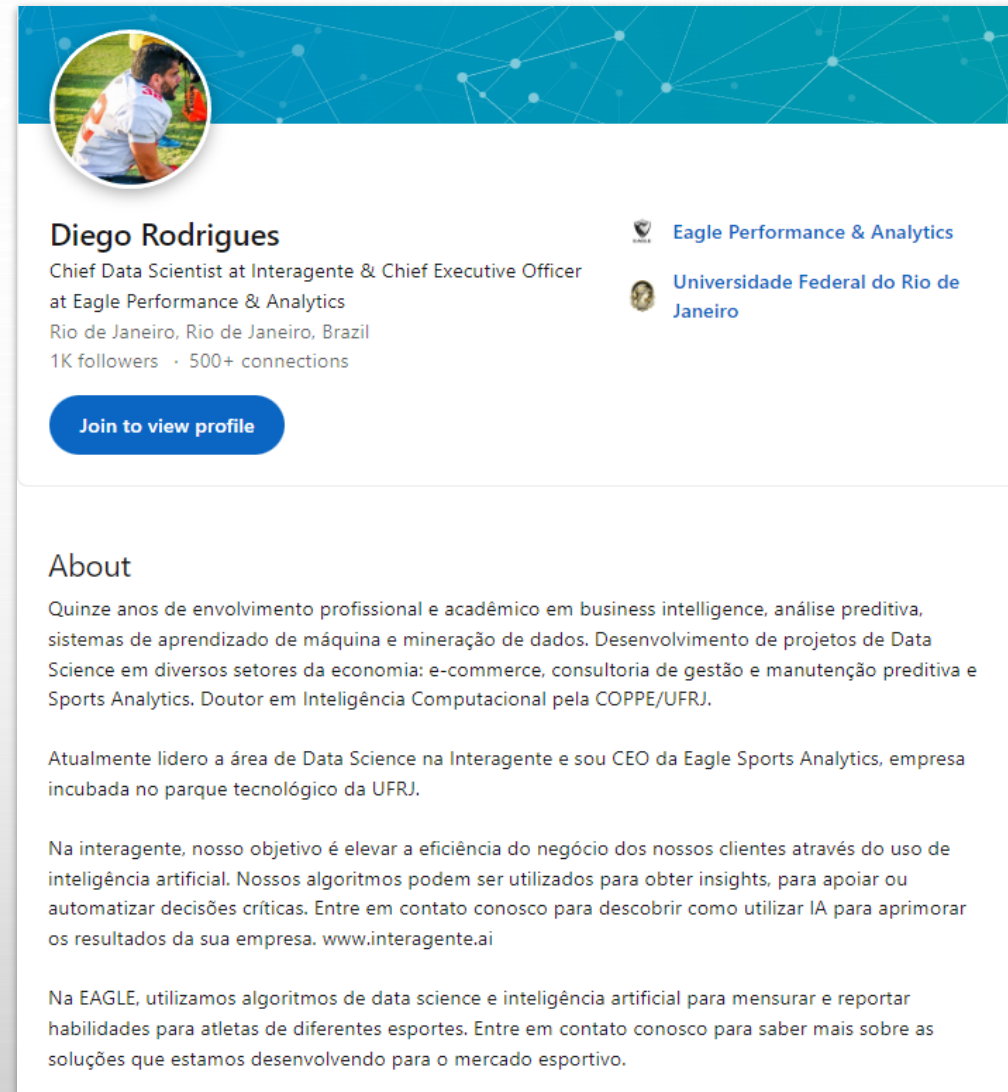
- INTRODUÇÃO
- PARADIGMAS DE APRENDIZADO DE MÁQUINA

PARTE 2 : PRÁTICA

- AMBIENTE DE DESENVOLVIMENTO
- BIBLIOTECAS

PROFESSOR DIEGO

- Doutor em Inteligência Computacional COPPE / UFRJ
- 17 anos de experiência com ciência de dados aplicada à engenharia.
- Sócio da Interagente.
- Professor da INFNET na Graduação, Pós-Graduação, Trilhas e Bootcamp.



The image shows a LinkedIn profile card for Diego Rodrigues. At the top is a circular profile picture of a man in a white shirt. Below the picture, the name 'Diego Rodrigues' is displayed in bold. Underneath the name, his current roles are listed: 'Chief Data Scientist at Interagente & Chief Executive Officer at Eagle Performance & Analytics'. His location is 'Rio de Janeiro, Rio de Janeiro, Brazil', and he has '1K followers' and '500+ connections'. To the right of the name, there are two logos: 'Eagle Performance & Analytics' and 'Universidade Federal do Rio de Janeiro'. A blue button with the text 'Join to view profile' is located below the name and location. Below the profile information, there is an 'About' section. The 'About' section contains two paragraphs of text. The first paragraph describes his 15 years of professional and academic experience in business intelligence, predictive analysis, machine learning systems, and data mining, as well as his work on Data Science projects in various sectors like e-commerce, consulting, and predictive maintenance and Sports Analytics. It also mentions his PhD in Computational Intelligence from COPPE/UFRJ. The second paragraph states that he currently leads the Data Science area at Interagente and is the CEO of Eagle Sports Analytics, a company incubated at UFRJ. It then describes the goal of Interagente, which is to improve business efficiency for clients using artificial intelligence, and mentions that their algorithms can be used to gain insights and support or automate critical decisions. It invites contact for more information on using AI to improve business results, with the website 'www.interagente.ai'. The third paragraph describes the work at EAGLE, which uses data science and artificial intelligence algorithms to measure and report athlete abilities for different sports, and also invites contact for more information on solutions being developed for the sports market.

Diego Rodrigues
Chief Data Scientist at Interagente & Chief Executive Officer
at Eagle Performance & Analytics
Rio de Janeiro, Rio de Janeiro, Brazil
1K followers · 500+ connections

[Join to view profile](#)

About

Quinze anos de envolvimento profissional e acadêmico em business intelligence, análise preditiva, sistemas de aprendizado de máquina e mineração de dados. Desenvolvimento de projetos de Data Science em diversos setores da economia: e-commerce, consultoria de gestão e manutenção preditiva e Sports Analytics. Doutor em Inteligência Computacional pela COPPE/UFRJ.

Atualmente lidero a área de Data Science na Interagente e sou CEO da Eagle Sports Analytics, empresa incubada no parque tecnológico da UFRJ.

Na interagente, nosso objetivo é elevar a eficiência do negócio dos nossos clientes através do uso de inteligência artificial. Nossos algoritmos podem ser utilizados para obter insights, para apoiar ou automatizar decisões críticas. Entre em contato conosco para descobrir como utilizar IA para aprimorar os resultados da sua empresa. www.interagente.ai

Na EAGLE, utilizamos algoritmos de data science e inteligência artificial para mensurar e reportar habilidades para atletas de diferentes esportes. Entre em contato conosco para saber mais sobre as soluções que estamos desenvolvendo para o mercado esportivo.

Produzir Ação

CICLO DE VIDA DO MODELO

Baseado em Dados



INTRODUÇÃO

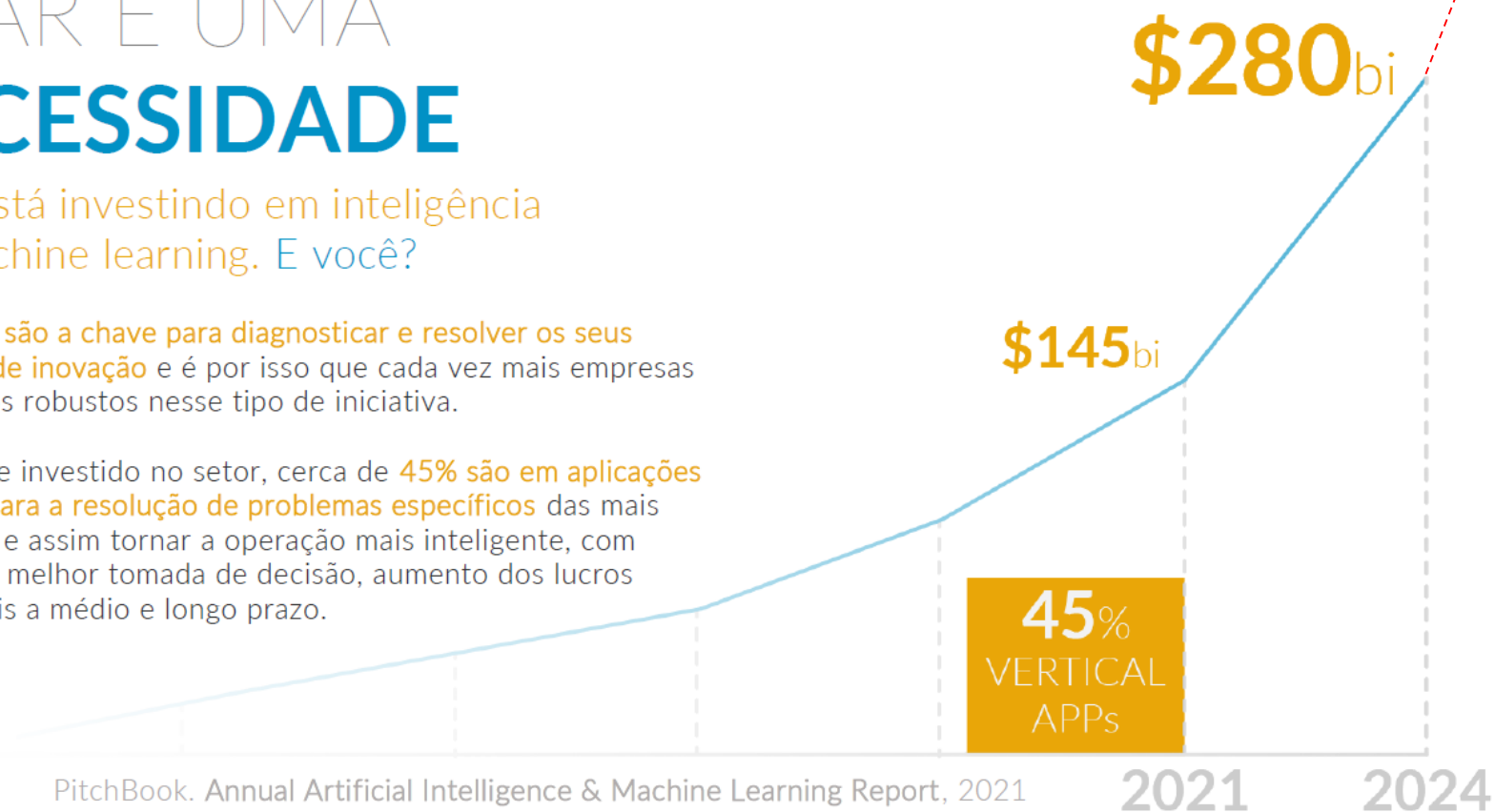
PROJEÇÃO 2024 DA ÁREA

INOVAR É UMA NECESSIDADE

O mundo já está investindo em inteligência artificial e machine learning. E você?

Os softwares de IA, são a chave para diagnosticar e resolver os seus principais desafios de inovação e é por isso que cada vez mais empresas fazem investimentos robustos nesse tipo de iniciativa.

De todo o montante investido no setor, cerca de 45% são em aplicações verticais, voltadas para a resolução de problemas específicos das mais variadas naturezas, e assim tornar a operação mais inteligente, com menos desperdício, melhor tomada de decisão, aumento dos lucros e resultados visíveis a médio e longo prazo.



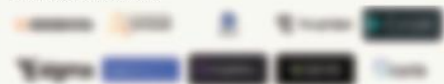
ECOSSISTEMA DE IA

AI & ML VC ecosystem market map

This market map is an overview of venture-backed or growth-stage companies that have received venture capital or other notable private investments. [Click to view the full map on the PitchBook Platform.](#)

1 Horizontal platforms

AI automation platforms



AI core



Computer vision



Foundation models



Natural language technology



2 Vertical applications

Consumer



Financial services



Healthcare



Industrial



IT



Transportation



3 Semiconductors

Graphics processing unit (GPU) cloud



Edge AI software



Intelligent sensors & devices



Processor design



4 Autonomous machines

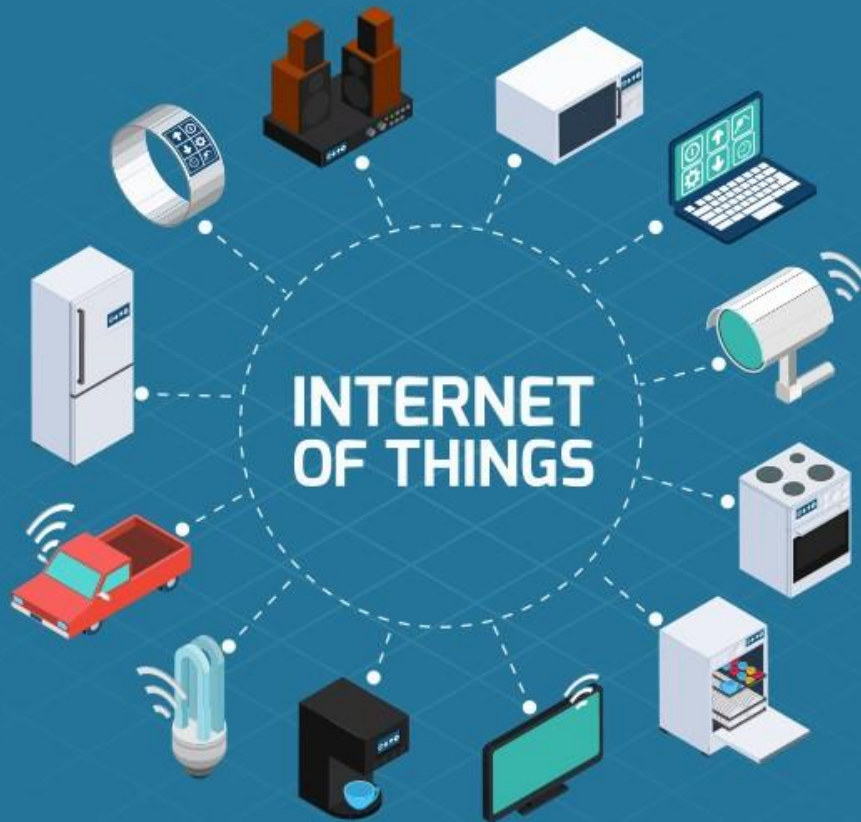
Autonomous vehicles



Intelligent robotics

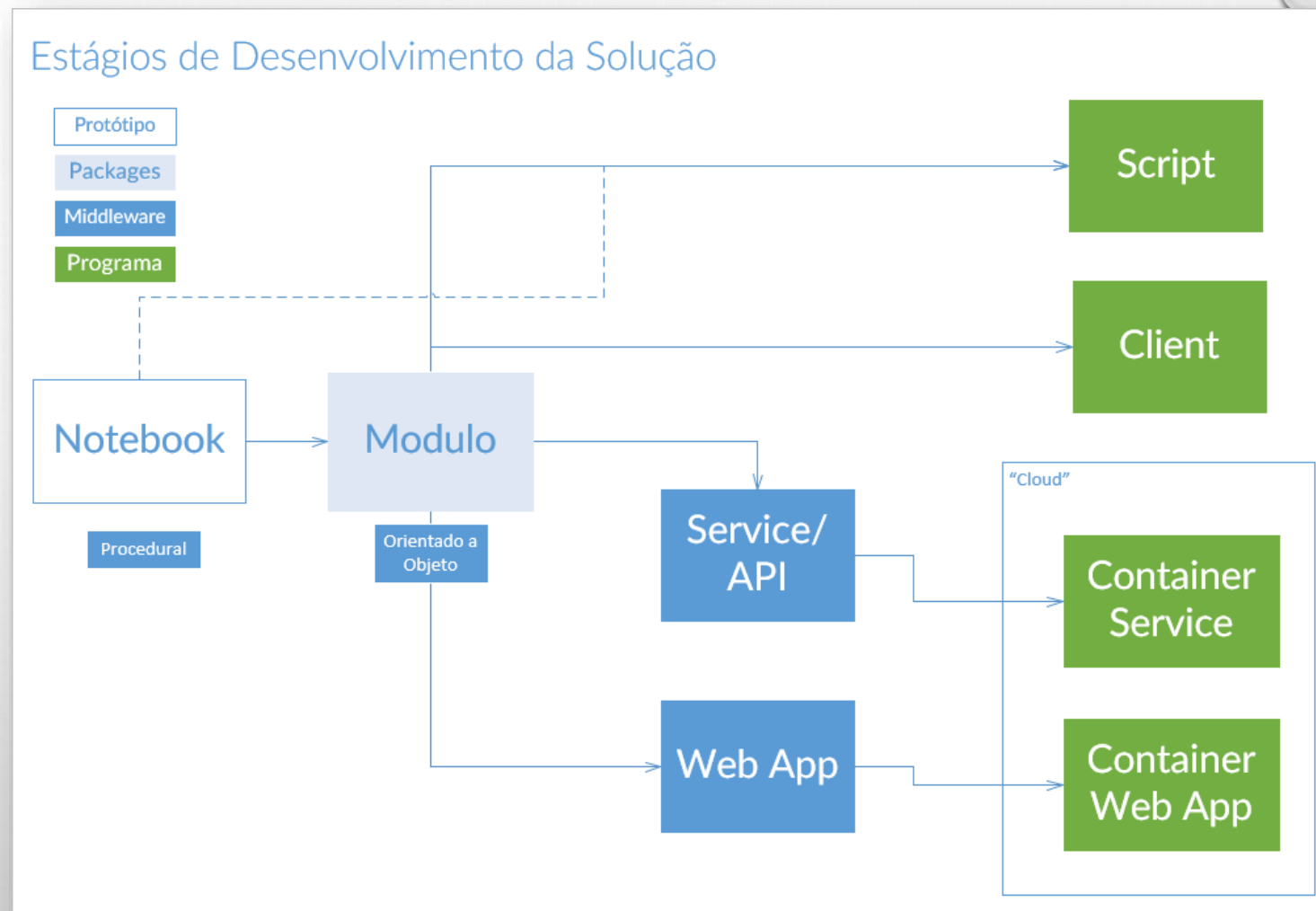


INTERNET DAS COISAS

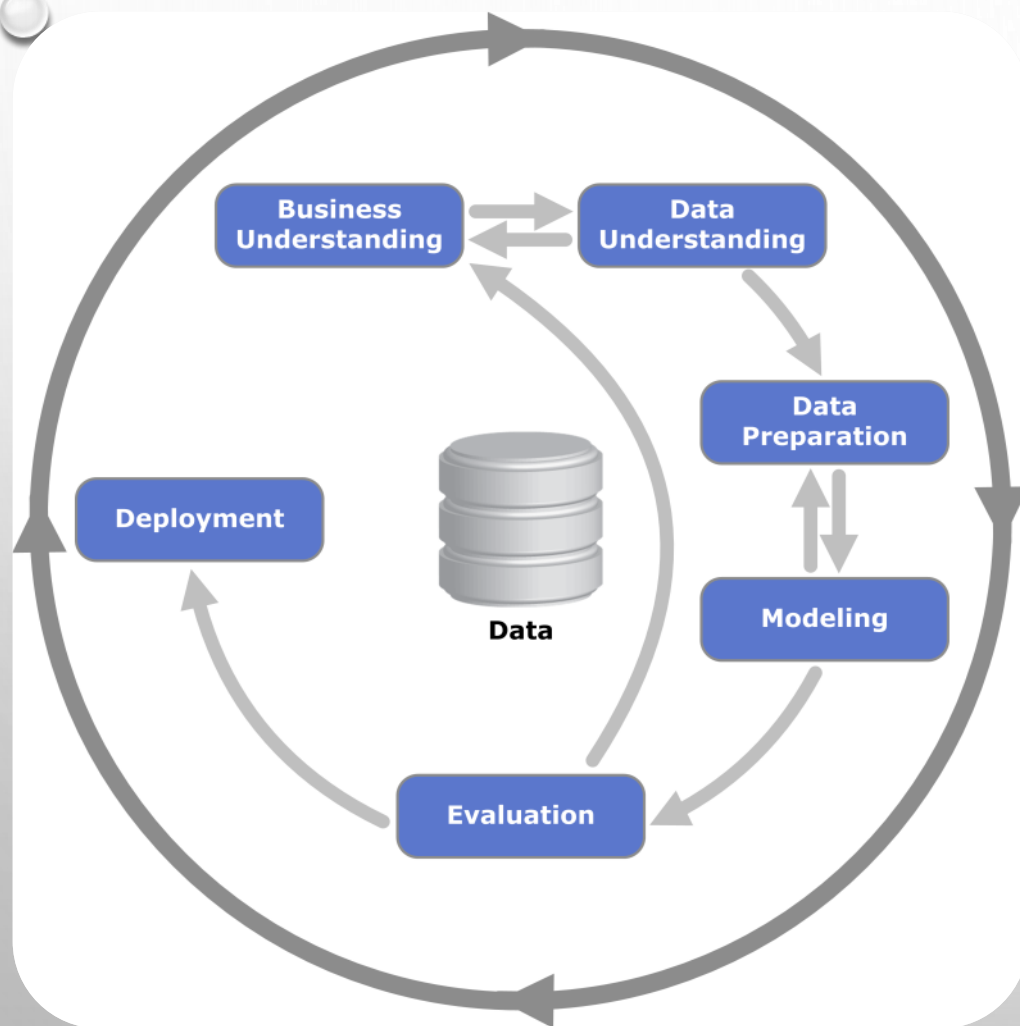


- Sensoriamento de Equipamentos
 - Previsão de Falha em Motor.
 - Previsão de Energia Eólica.
- Web Scrapping
 - Análise de Sentimento Netflix
- APIs
 - Classificação de Notícias
 - Classificação de Raças de Cachorros
- Dispositivos Móveis
 - Censor de Gestos Agressivos
 - Detecção de Manobras Agressivas

APLICAÇÕES



Cross Industry Standard Process for Data Mining - IBM



1) **Requerimentos e Análise de Negócio**

Entendimento do problema decisório, dados relacionados & revisão bibliográfica.

2) **Preparação dos Dados**

Entendimento das fontes de dados, dos tipos e elaboração da representação.

3) **Modelagem**

Análise Exploratória, Seleção de atributos e treinamento.

4) **Avaliação**

Seleção do melhor modelo.

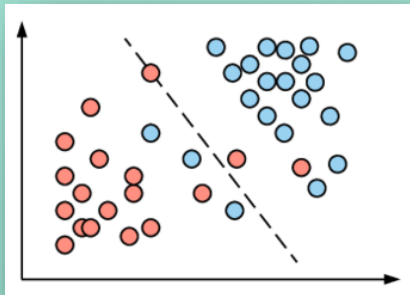
5) **Liberação**

Liberação do modelo no ambiente de produção.

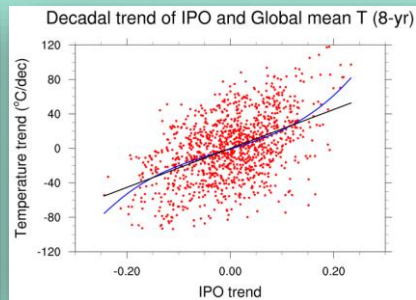
The background is a light gray gradient. In the top-left and bottom-right corners, there are several realistic water droplets of various sizes, some overlapping. A faint, circular watermark is visible in the upper center of the slide.

PARADIGMAS DE APRENDIZADO DE MÁQUINA

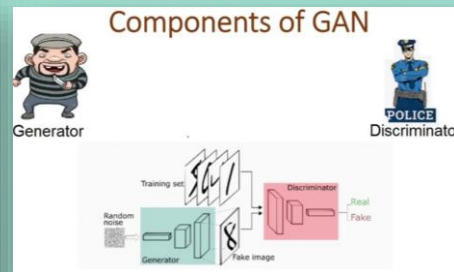
APRENDIZADO SUPERVISIONADO



CLASSIFICAÇÃO

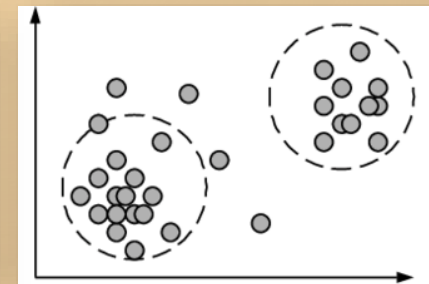


REGRESSÃO



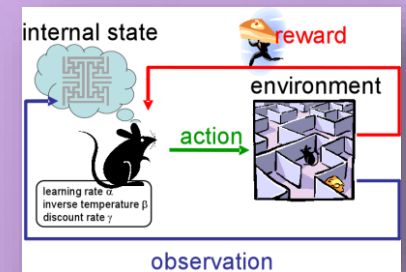
GENERATIVO

APRENDIZADO NÃO- SUPERVISIONADO



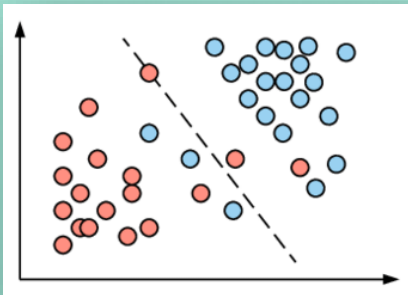
AGRUPAMENTO

APRENDIZADO POR REFORÇO

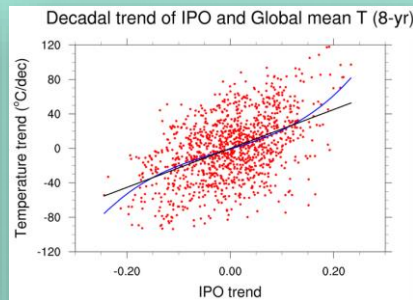


REFORÇO

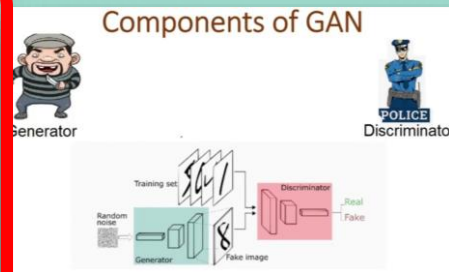
APRENDIZADO SUPERVISIONADO



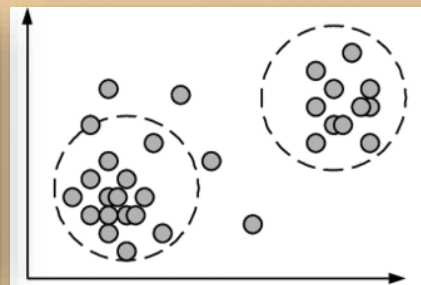
CLASSIFICAÇÃO



REGRESSÃO



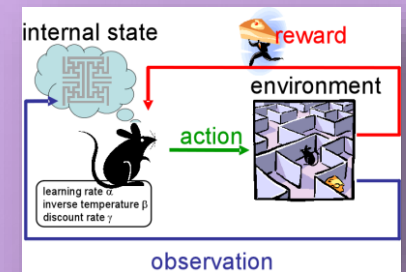
GENERATIVO



AGRUPAMENTO

APRENDIZADO NÃO- SUPERVISIONADO

APRENDIZADO POR REFORÇO



REFORÇO

APRENDIZADO SUPERVISIONADO

Tarefas de **classificação** e **regressão** pertencem a esta categoria. O treinamento consiste em **encontrar parâmetros** para o modelo que **minimiza uma função de risco/erro** para uma amostra de treinamento, baseado na diferença entre os **valores previstos e reais**, para cada observação.

CLASSIFICAÇÃO

Um bebê consegue separar e ordenar blocos com diferentes tamanhos, formas e cores. Ele também consegue **identificar os tipos diferentes de objetos**.

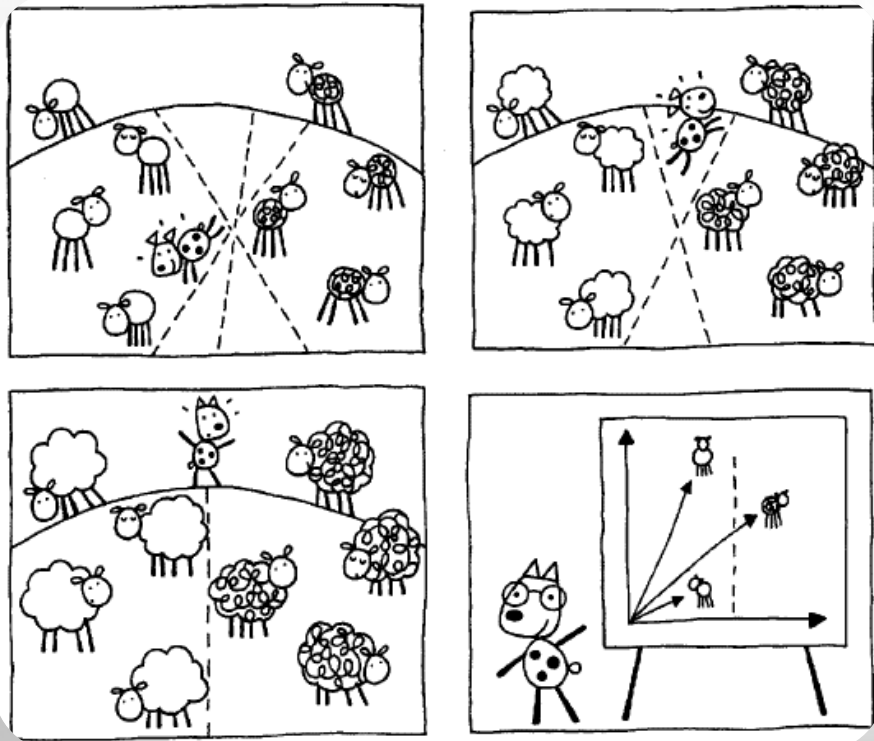
Os diferentes tipos de objetos são chamados de **classes**. As características dos objetos são chamadas de **variáveis** ou **atributos**.



Data Scientist da Nova Geração

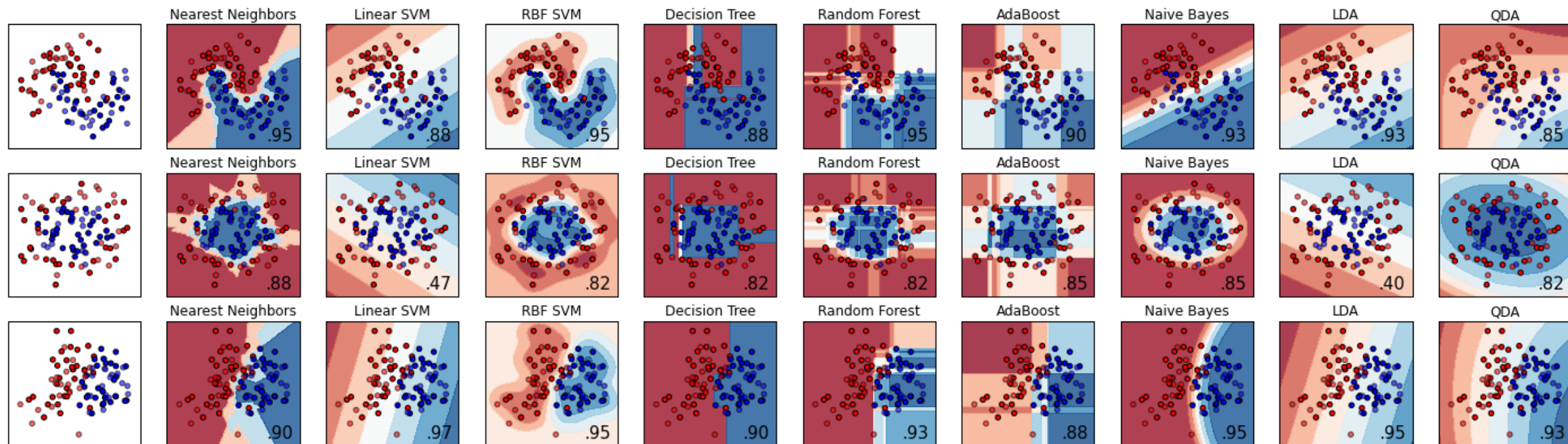
Então, um classificador é um modelo **treinado para discriminar objetos** pertencentes a duas ou mais classes, baseado em seus atributos.

REPRESENTAÇÃO



Ideia: como quantificar um objeto no mundo físico no mundo digital?

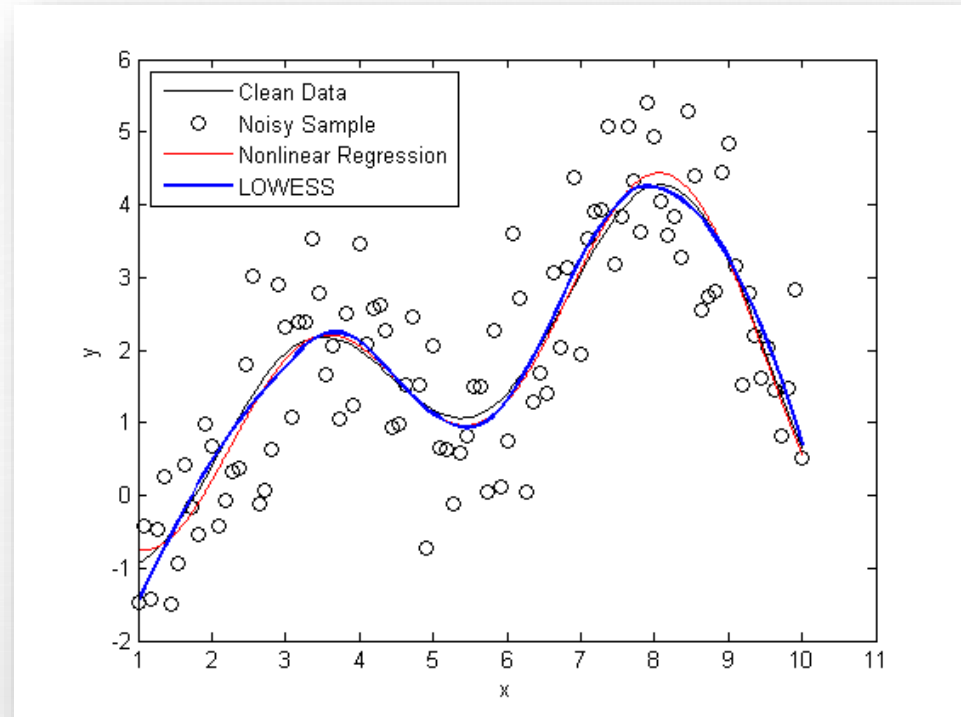
Exercício: qual seria uma boa representação para diferenciar ratos e elefantes?



REGRESSÃO

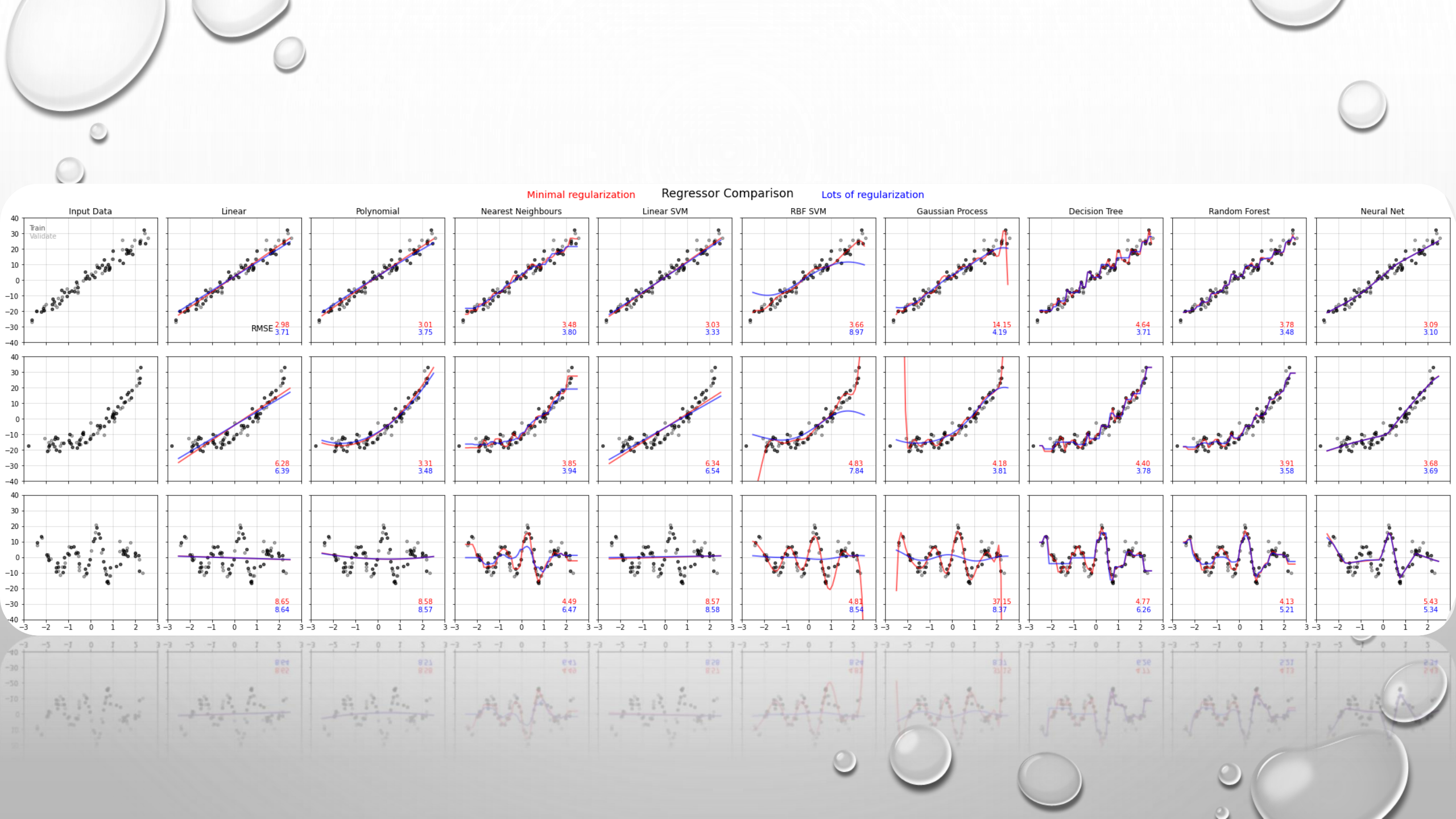
O objetivo da regressão é
modelar as relações funcionais
entre dois conjuntos de variáveis.

As variáveis que representam as causas são chamadas de **variáveis independentes**, e as variáveis cujo objetivo é prever, são chamadas **variáveis dependentes**.

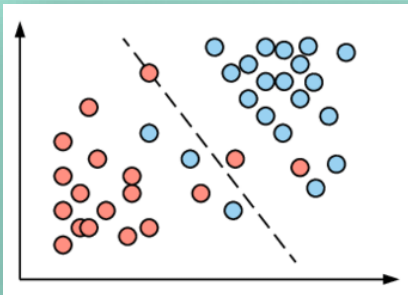


As vezes quando o mundo
não é linear & gaussiano...

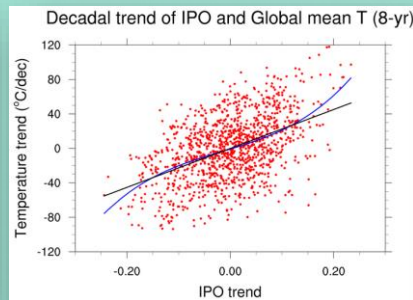
Então, uma **regressão** é um modelo utilizado para
prever **uma ou mais variáveis dependentes**,
baseado em causas, ou variáveis independentes.



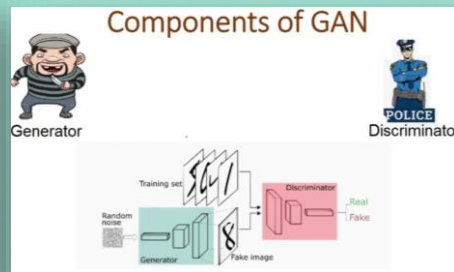
APRENDIZADO SUPERVISIONADO



CLASSIFICAÇÃO

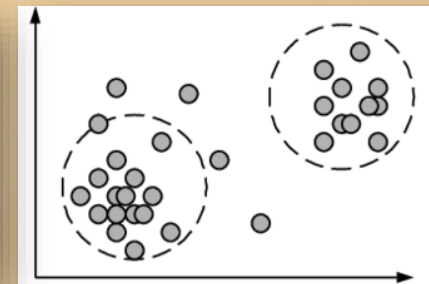


REGRESSÃO



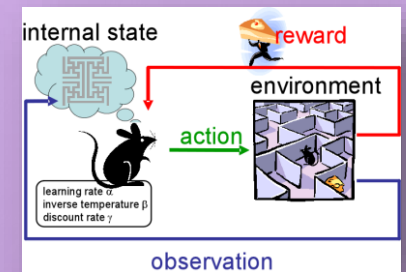
GENERATIVO

APRENDIZADO NÃO- SUPERVISIONADO



AGRUPAMENTO

APRENDIZADO POR REFORÇO




REFORÇO



○ Aprendizado Não-Supervisionado

Não existe um **conhecimento “a priori” dos grupos** contidos nos dados. Algoritmos de agrupamento dependem fortemente de uma definição de **“distância”** ou **“similaridade”** ou **“probabilidade”** entre as observações.



Agrupamento

Um bebê consegue **agrupar objetos por cor, tamanho, formato** e muitos outros atributos que ele pode observar nos objetos.

Diferentes maneiras de organizar os objetos são diferentes **estruturas de agrupamentos** existentes em uma amostra de dados.



De quantas maneiras estes blocos podem ser organizados em grupos?

Um **modelo de agrupamento** é usado para **identificar grupos**, ou estruturas de agrupamentos, nos dados.

Modelagem Probabilística

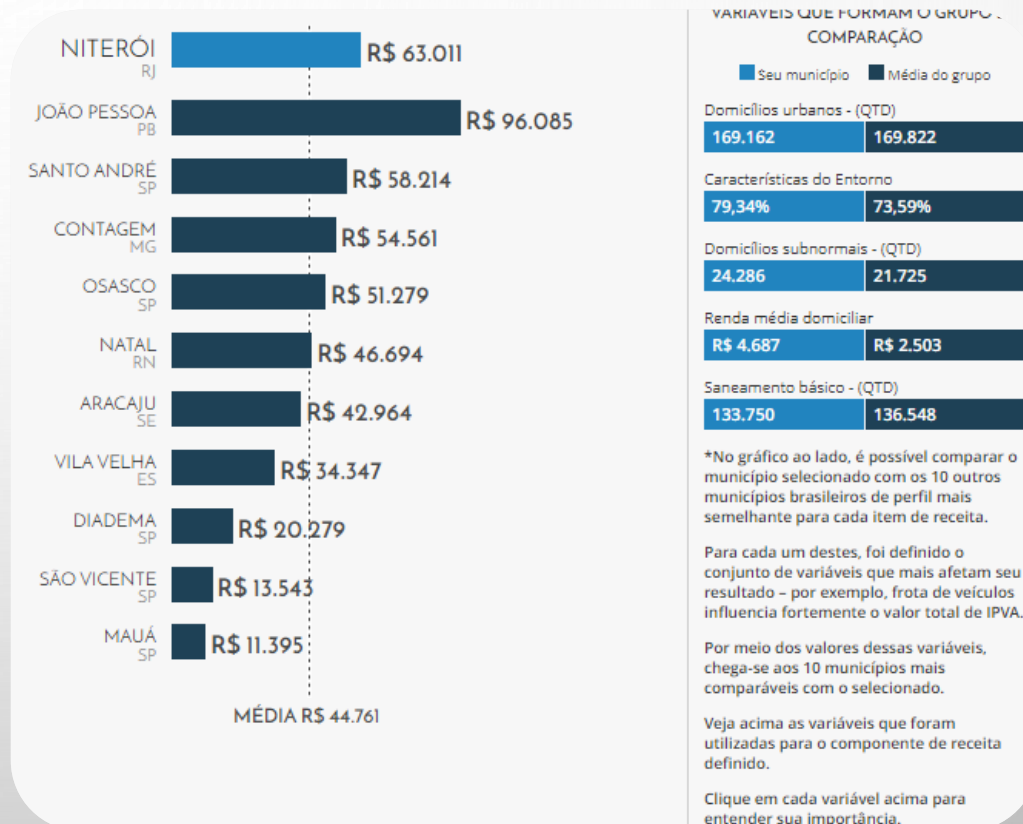
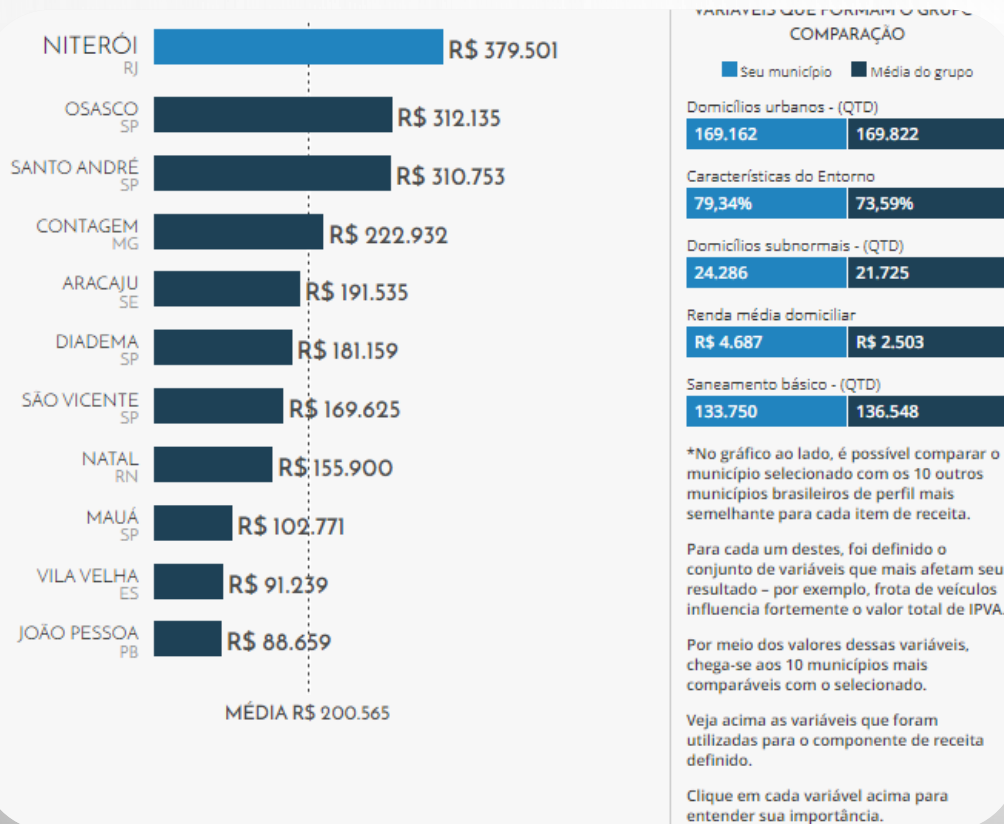
Em uma cesta de supermercado existe uma variedade grande de itens comprados juntos. A presença de um item específico afeta a **chance de outro produto estar presente na cesta?**



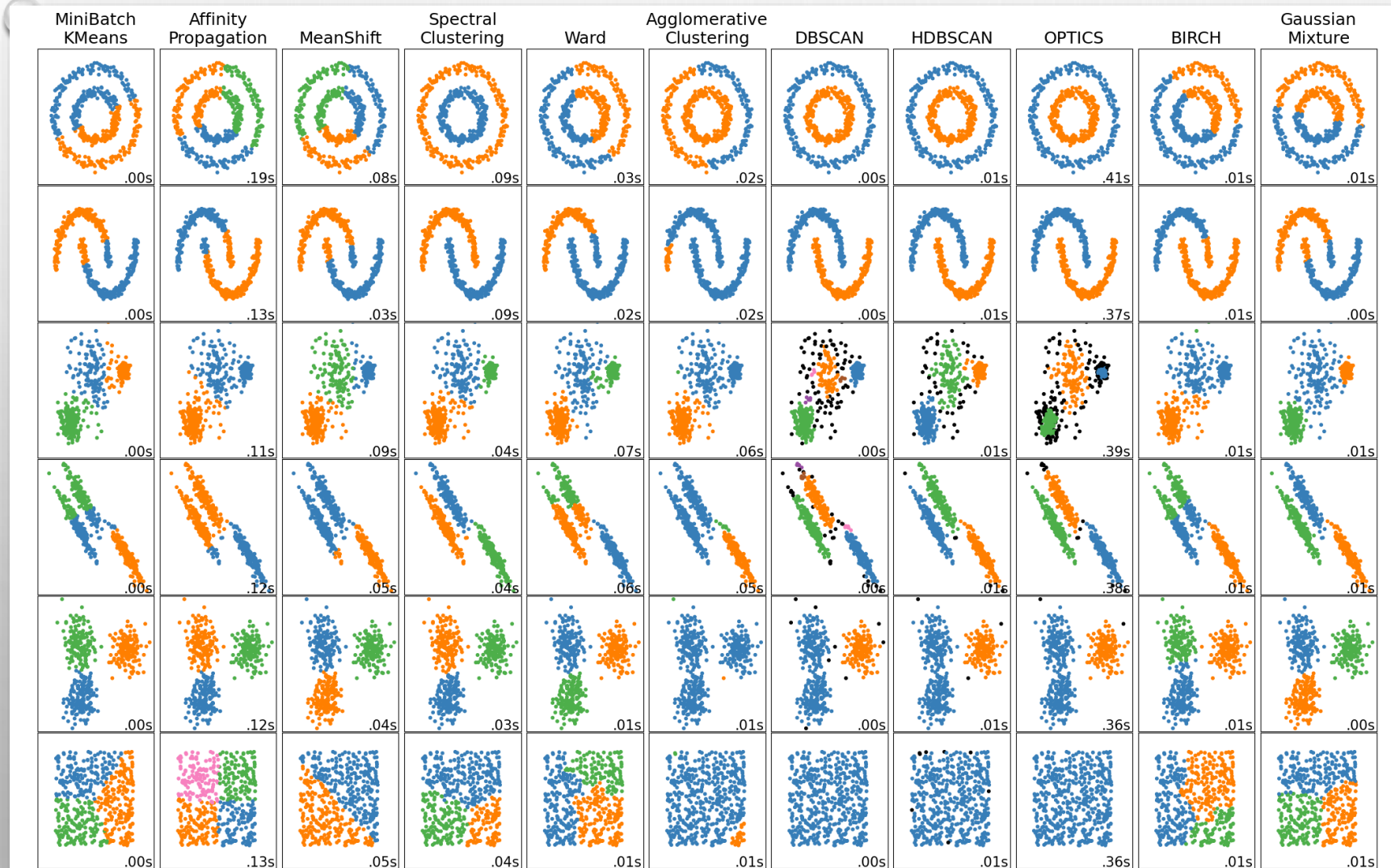
Será que algum desses itens são comprados juntos?

O objetivo de modelos probabilísticos é **identificar padrões** dentro de comportamentos supostamente **aleatórios**.

REPRESENTAÇÃO: COMO ENCONTRAR OS 10 MUNICÍPIOS MAIS SIMILARES A NITERÓI?



MODELAGEM



QUANTOS
AGRUPAMENTOS
EXISTEM NESSES
DADOS?

MODELAGEM

Topic Word Scores

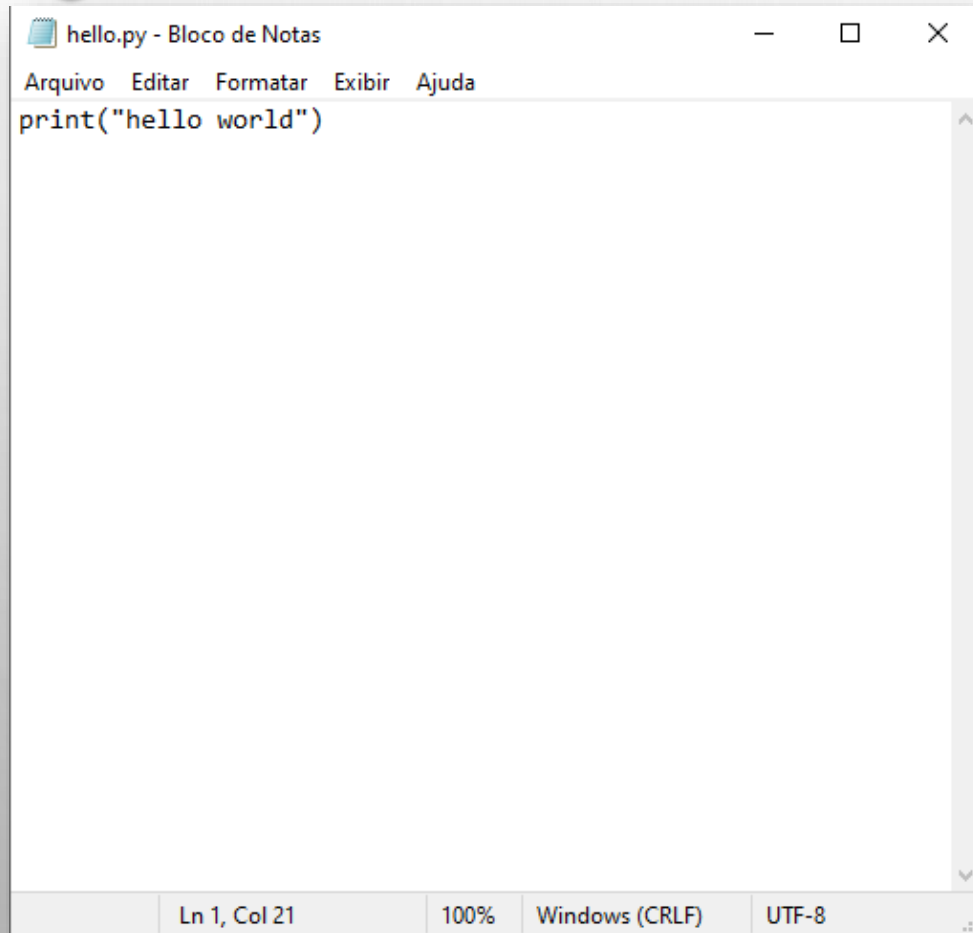


QUAIS TÓPICOS
QUENTES EM REDES
SOCIAIS NERDS?

The background of the slide is a light gray gradient. In the top-left and bottom-right corners, there are several realistic water droplets of various sizes, rendered with soft shadows and highlights to give them a three-dimensional appearance. In the center of the slide, there is a faint, circular watermark. It features a globe with latitude and longitude lines, and the text 'BRASIL' is visible at the bottom of the circle.

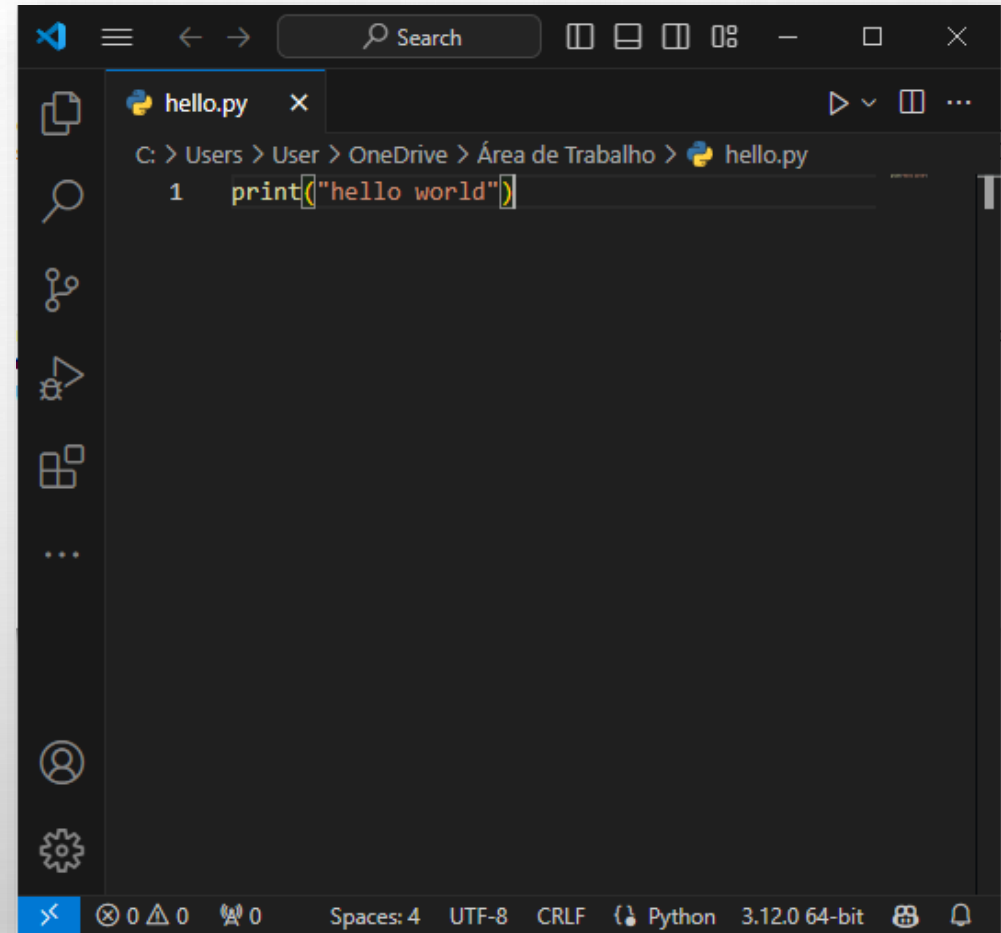
AMBIENTE DE DESENVOLVIMENTO

EDITOR DE TEXTO OU IDE (INTEGRATED DEVELOPMENT ENVIROMENT)?



A screenshot of a basic text editor window titled "hello.py - Bloco de Notas". The window has a menu bar with "Arquivo", "Editar", "Formatar", "Exibir", and "Ajuda". The main text area contains the code `print("hello world")`. The status bar at the bottom shows "Ln 1, Col 21", "100%", "Windows (CRLF)", and "UTF-8".

Amadores...



A screenshot of a modern IDE window titled "hello.py". The window has a dark theme and a sidebar on the left with icons for Explorer, Search, Source Control, Run and Debug, Extensions, and Settings. The main text area shows the code `1 print("hello world")` with line numbers. The status bar at the bottom displays "0 0 0", "Spaces: 4", "UTF-8", "CRLF", "Python 3.12.0 64-bit", and a Python icon.

PROS!

AMBIENTE DE DESENVOLVIMENTO

Miniconda

Miniconda is a free minimal installer for conda. It is a small bootstrap version of Anaconda that includes only conda, Python, the packages they both depend on, and a small number of other useful packages (like pip, zlib, and a few others). If you need more packages, use the `conda install` command to install from thousands of packages available by default in Anaconda's public repo, or from other channels, like conda-forge or bioconda.

Is Miniconda the right conda install for you? The [Anaconda or Miniconda](#) page lists some reasons why you might want one installation over the other.

[System requirements](#)

[Latest Miniconda installer links by Python version](#)

[Installing Miniconda](#)

[Miniconda release notes](#)

[Other resources](#)

[Miniconda hash information](#)

Simple Python Version Management: pyenv

[gitter](#) [join chat](#)

pyenv lets you easily switch between multiple versions of Python. It's simple, unobtrusive, and follows the UNIX tradition of single-purpose tools that do one thing well.

This project was forked from [rbenv](#) and [ruby-build](#), and modified for Python.

```
$ pyenv versions
2.7.10
* 3.5.0 (set by /Users/yuu/.pyenv/version)
miniconda3-3.16.0
pypy-2.6.0

$ python --version
Python 3.5.0

$ pyenv global pypy-2.6.0

$ python --version
Python 2.7.9 (295ee98b69288471b0fcf2e0ede82ce5209eb90b, Jun 01 2015, 17:30:13)
[PyPy 2.6.0 with GCC 4.9.2]

$ cd /Volumes/treasuredata/jupyter

$ pyenv version
miniconda3-3.16.0 (set by /Volumes/treasuredata/.python-version)

$ python --version
Python 3.4.3 :: Continuum Analytics, Inc.
```

Boa Prática: 1 Ambiente por projeto!

CRIANDO AMBIENTE COM CONDA

Managing environments

With conda, you can create, export, list, remove, and update environments that have different versions of Python and/or packages installed in them. Switching or moving between environments is called activating the environment. You can also share an environment file.

There are many options available for the commands described on this page. For a detailed reference on all available commands, see [commands](#).

Creating an environment with commands

Use the terminal for the following steps:

1. To create an environment:

```
conda create --name <my-env>
```

Replace `<my-env>` with the name of your environment.

2. When conda asks you to proceed, type `y`:

```
proceed ([y]/n)?
```

This creates the myenv environment in `/envs/`. No packages will be installed in this environment.

3. To create an environment with a specific version of Python:

```
conda create -n myenv python=3.9
```

Activating an environment

Activating environments is essential to making the software in the environments work well. Activation entails two primary functions: adding entries to PATH for the environment and running any activation scripts that the environment may contain. These activation scripts are how packages can set arbitrary environment variables that may be necessary for their operation. You can also [use the config API to set environment variables](#).

Activation prepends to PATH. This only takes effect when you have the environment active so it is local to a terminal session, not global.

Note

When [installing Anaconda](#), you have the option to "Add Anaconda to my PATH environment variable." *This is not recommended* because it *appends* Anaconda to PATH. When the installer appends to PATH, it does not call the activation scripts.

Note

On Windows, PATH is composed of two parts, the *system* PATH and the *user* PATH. The system PATH always comes first. When you install Anaconda for "Just Me", we add it to the *user* PATH. When you install for "All Users", we add it to the *system* PATH. In the former case, you can end up with system PATH values taking precedence over your entries. In the latter case, you do not. *We do not recommend* [multi-user installs](#).

To activate an environment: `conda activate myenv`



BIBLIOTECAS



AMBIENTE PYTHON



6. Machine Learning

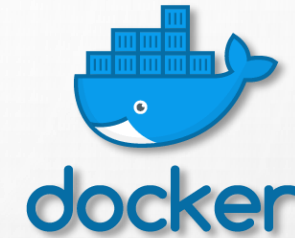


6. Deployment



4. Variáveis Aleatórias

5. Visualização



1. Editor de Código



2. Gestor de Ambiente



3. Ambiente Python do Projeto



3. Notebook Dinâmico

The background is a light gray gradient. In the top-left and bottom-right corners, there are several realistic water droplets of various sizes, rendered with soft shadows and highlights to give them a three-dimensional appearance. The text is centered in the middle of the slide.

CONCEITOS DE MACHINE LEARNING II