

Introdução ao Desenvolvimento de Modelos de Aprendizado de Máquina Produtivos

Moacir Antonelli Ponti

ICMC/USP
Mercado Livre

Wecomp 2024, São Carlos-SP, Brasil

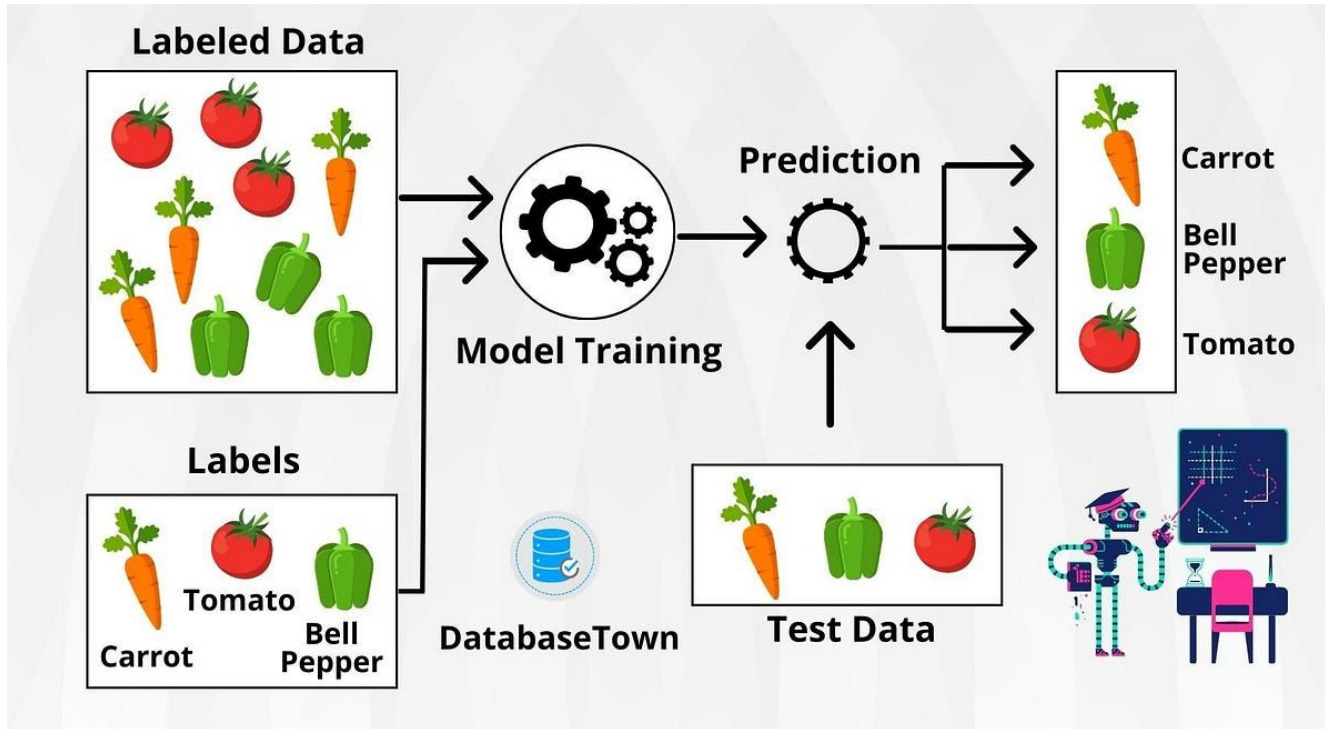
Disclaimer

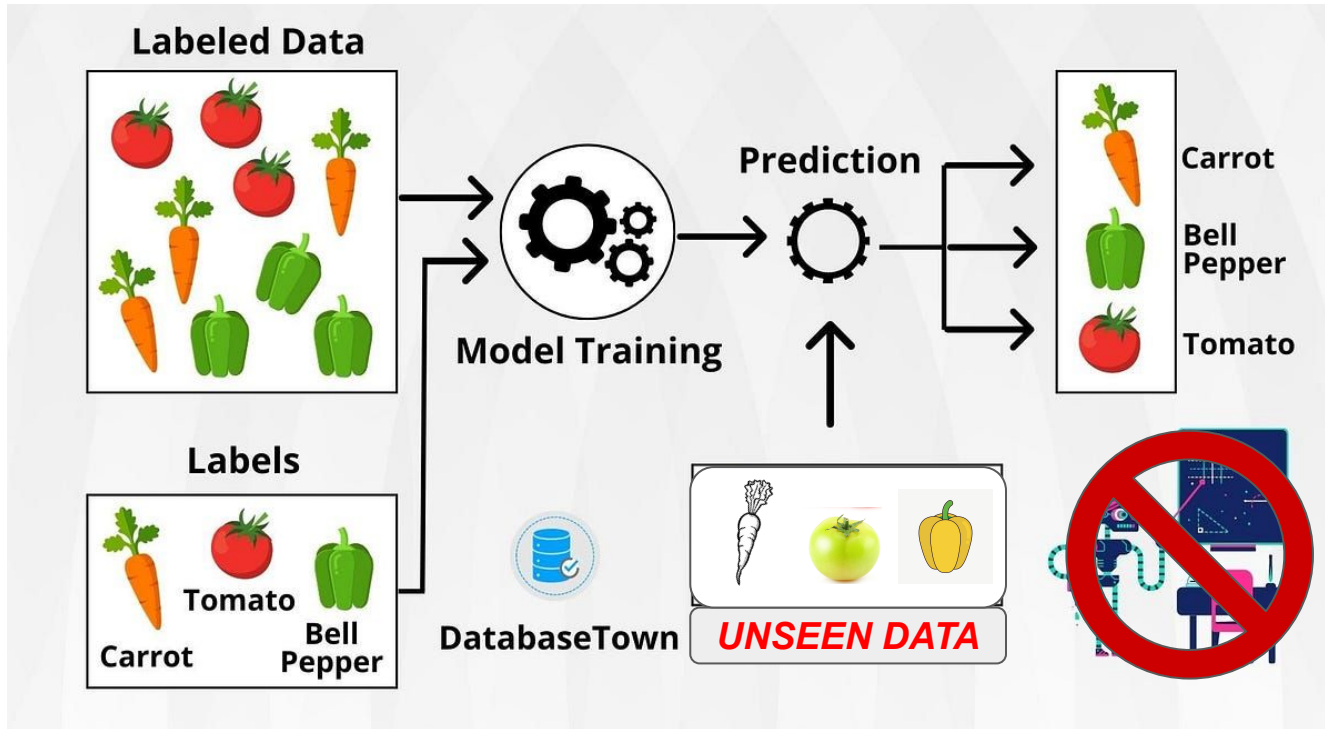
- Conteúdo e opiniões não representam a visão do Mercado Livre sobre o assunto.

Aprendizado Supervisionado

$$X, Y = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}, \text{ for } x_i \in \mathbb{R}^M$$

$$f : X \rightarrow Y$$





A criação de sistemas de aprendizado de máquina produtivos, envolve uma série de preocupações não encontradas em *exemplos tutoriais* ou mesmo em *grandes experimentos de pesquisa offline*

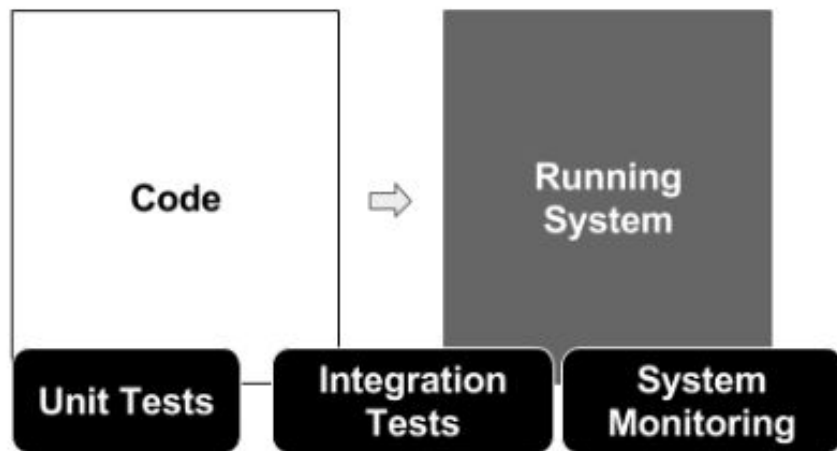
Eric Breck, Shanqing Cai, Eric Nielsen, Michael Salib, D. Sculley (Google)
"The ML Test Score" / IEEE Big Data, 2017

**The ML Test Score:
A Rubric for ML Production Readiness and Technical Debt Reduction**

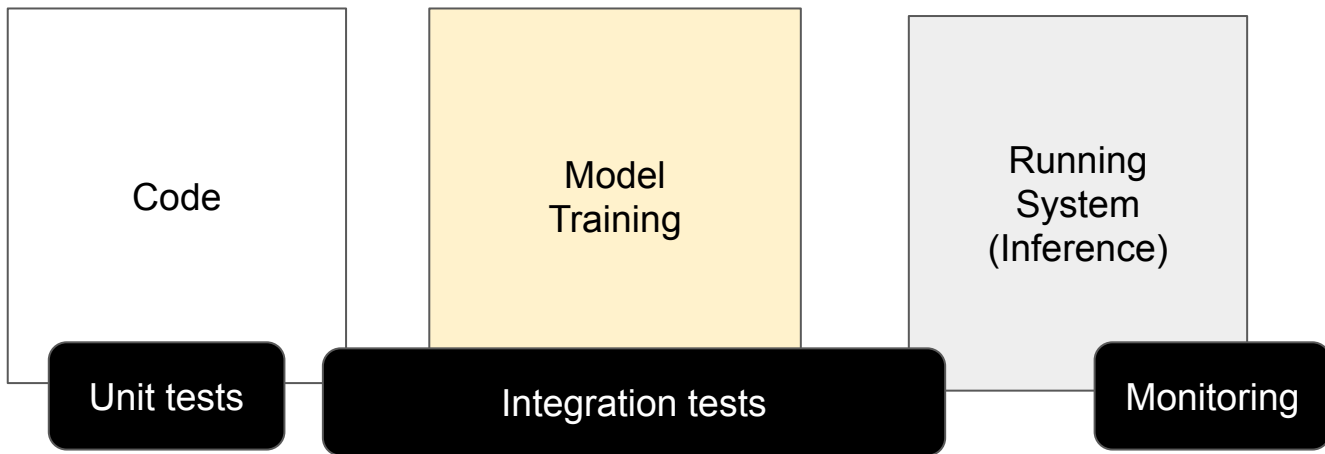
Eric Breck, Shanqing Cai, Eric Nielsen, Michael Salib, D. Sculley
Google, Inc.
ebreck, cais, nielsene, msalib, dsculley@google.com

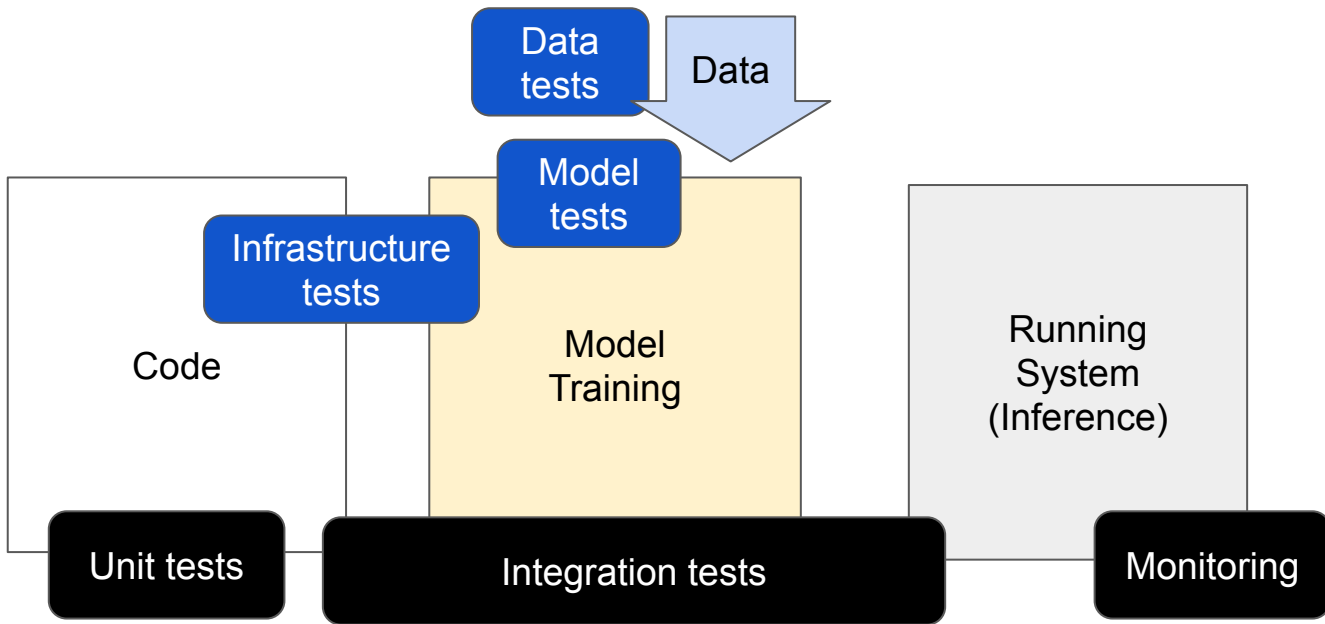
Abstract—Creating reliable, production-level machine learning systems brings on a host of concerns not found in small toy examples or even large offline research experiments. Testing and monitoring are key considerations for ensuring the production-readiness of an ML system, and for reducing technical debt of ML systems. But it can be difficult to formu-

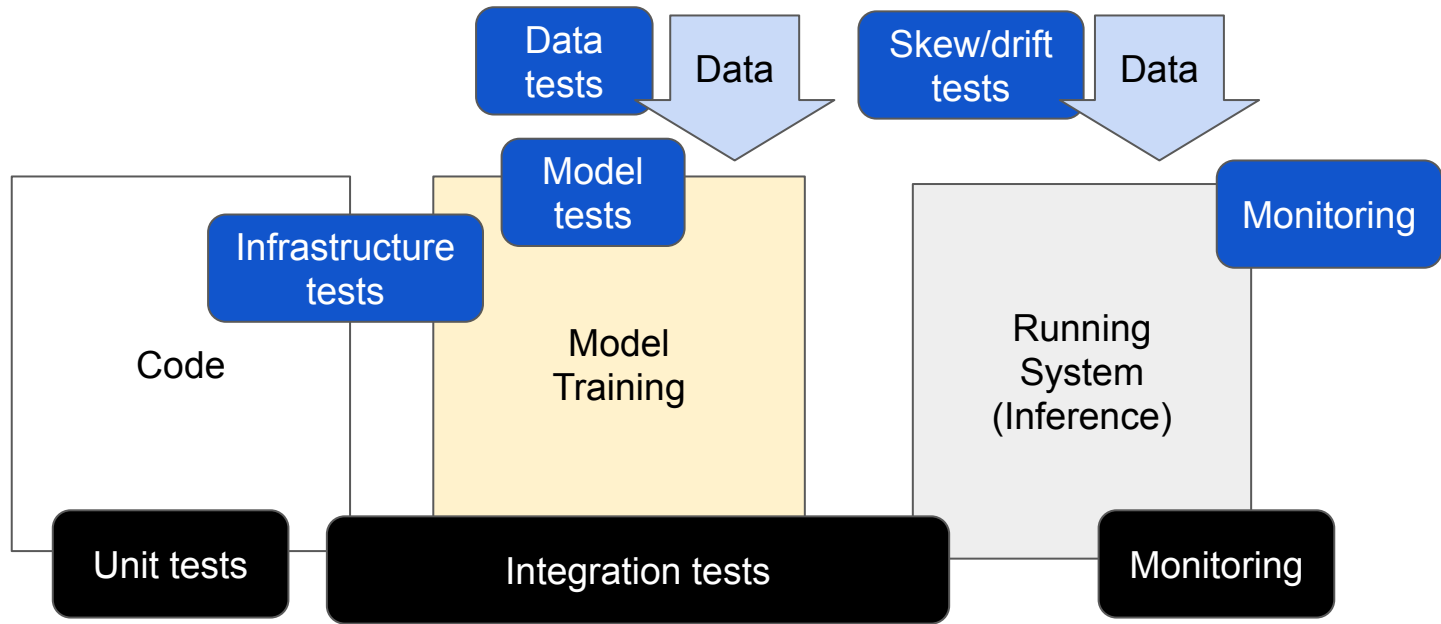
may find difficult. Note that this rubric focuses on issues specific to ML systems, and so does not include generic software engineering best practices such as ensuring good unit test coverage and a well-defined binary release process. Such strategies remain necessary as well. We do call out



Traditional System Testing and Monitoring

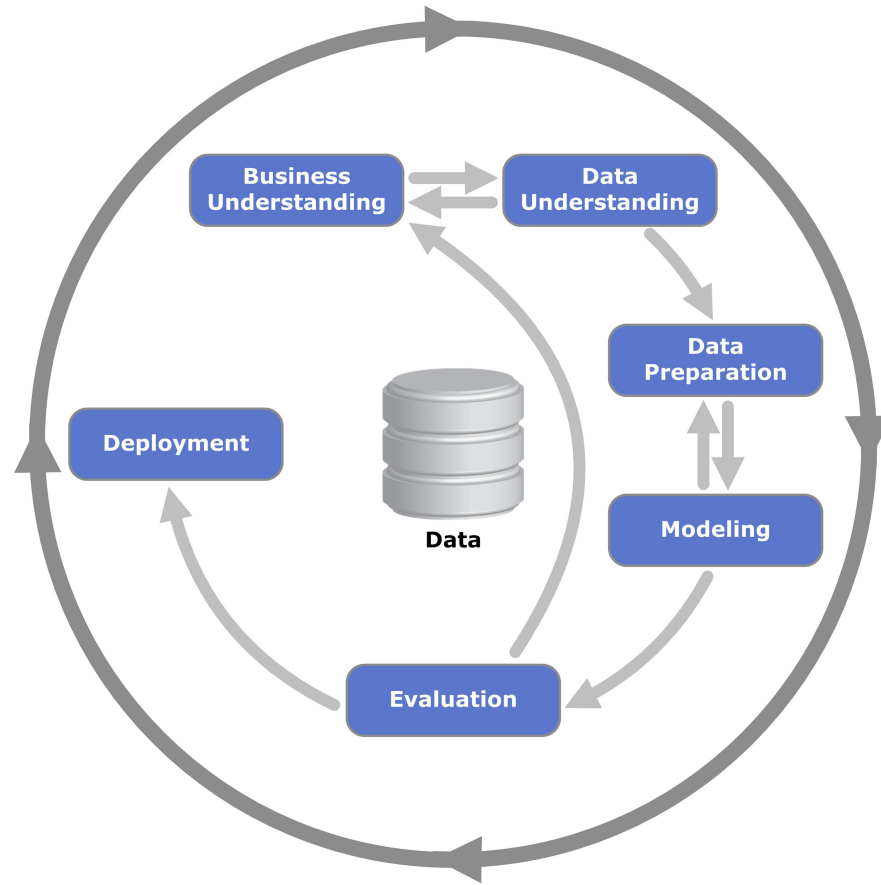






Agenda

1. Metodologia de desenvolvimento para aprendizado de máquina/ciência de dados
2. Dataset/ETL (Extract, Transform, Load)
3. Treinamento de um modelo
4. Modelo em inferência



CRISP DM

Cross Industry Standard Process for Data Mining

Aplicação real: moderação de itens proibidos

[Salud y Equipamiento Médico](#) > [Equipamiento Médico](#) > [Otros](#)

Equipo De Rayos X Odontológico

Medicamentos

\$ 39.999,00 ARS / \$ 419.41 USD

Sin precio de referencia \$ 43.998,90 ARS (+10%)

Stock: 1

Core

Usado

Gratuita



Beleza e Cuidado Pessoal > Higiene Pessoal > Desodorantes

Promoção Limitada 1 Frasco Pronto Entrega

R\$ 90,00 BRL / \$ 15.70 USD

R\$ 0,00 BRL - R\$ 2,23 BRL R\$ 99,00 BRL (+10%)

Stock: 19

1 unidade

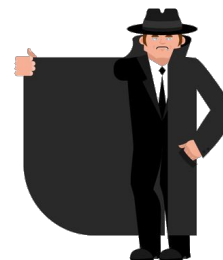




**Respeitar leis e
regulamentações**



**Melhorar a segurança de
compradores**



**Identificar e reportar
usuários nocivos**

Frete grátis



 **COMPRA INTERNACIONAL**

Frete grátis do mundo até
sua casa



Lojas oficiais

Somente lojas oficiais (215)

Categorias

Materiais Escolares (170)

Tablets e Acessórios (120)

Artes e Atividades (51)

Periféricos para PC (246)

Brinquedos Eletrônicos (101)

Celulares e Telefones (26)

Antiguidades e Coleções (24)

Portáteis e Acessórios (24)

Bebês (16)

[Mostrar mais](#)



Tablet De Escrita Lcd Almofada De Desenho De 10,5 Polegadas,

R\$ 112³⁸

em 6x R\$ 18⁷³ sem juros

Frete grátis

COMPRA INTERNACIONAL



H12 12 12 Polegadas Lcd Escrita Digital Desenho Tablet Handw

~~R\$ 45³⁹~~

R\$ 99²⁰ 34% OFF

em 6x R\$ 16⁵³ sem juros

Frete grátis

COMPRA INTERNACIONAL



Tablet De Escrita Lcd Recarregável Desenho De Escrita Manual

Vendido por Top Store

R\$ 246⁵⁶

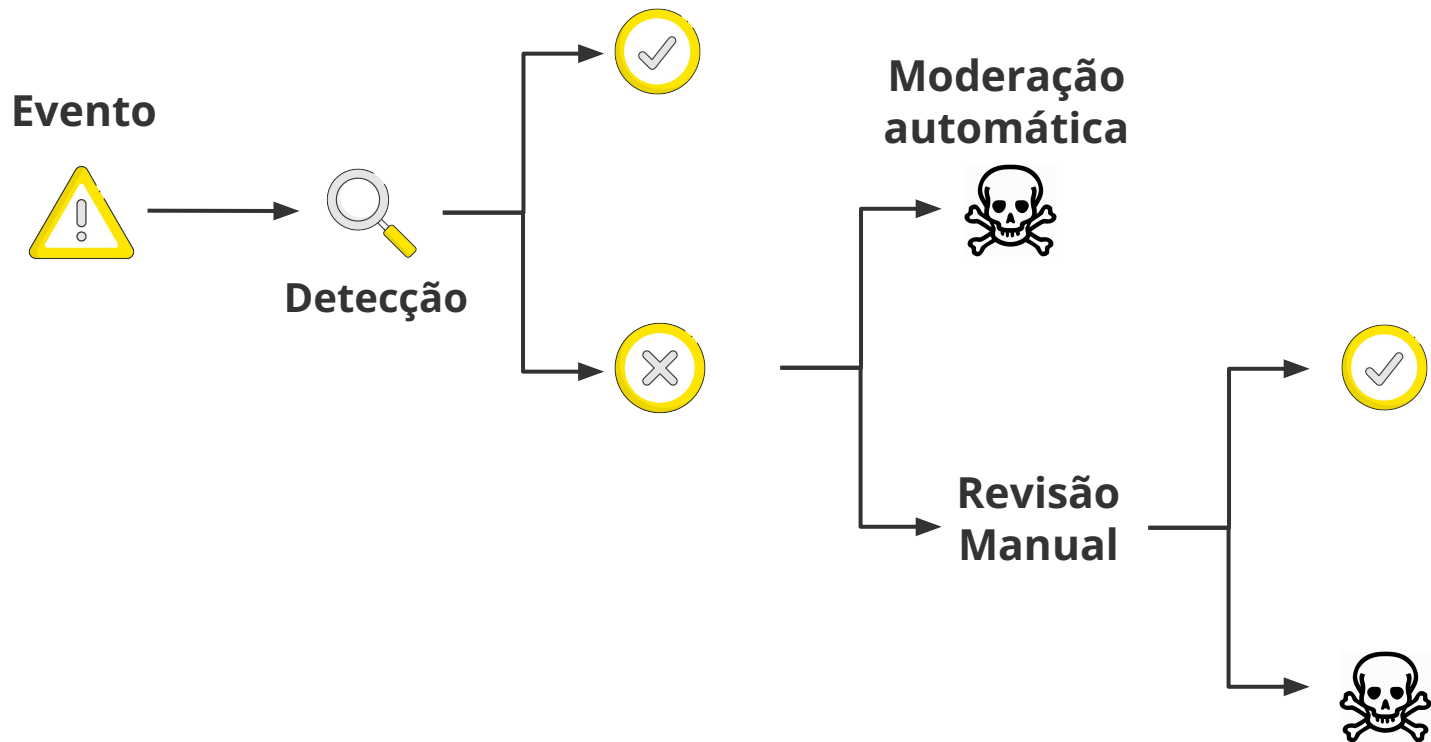
em 8x R\$ 30⁸² sem juros

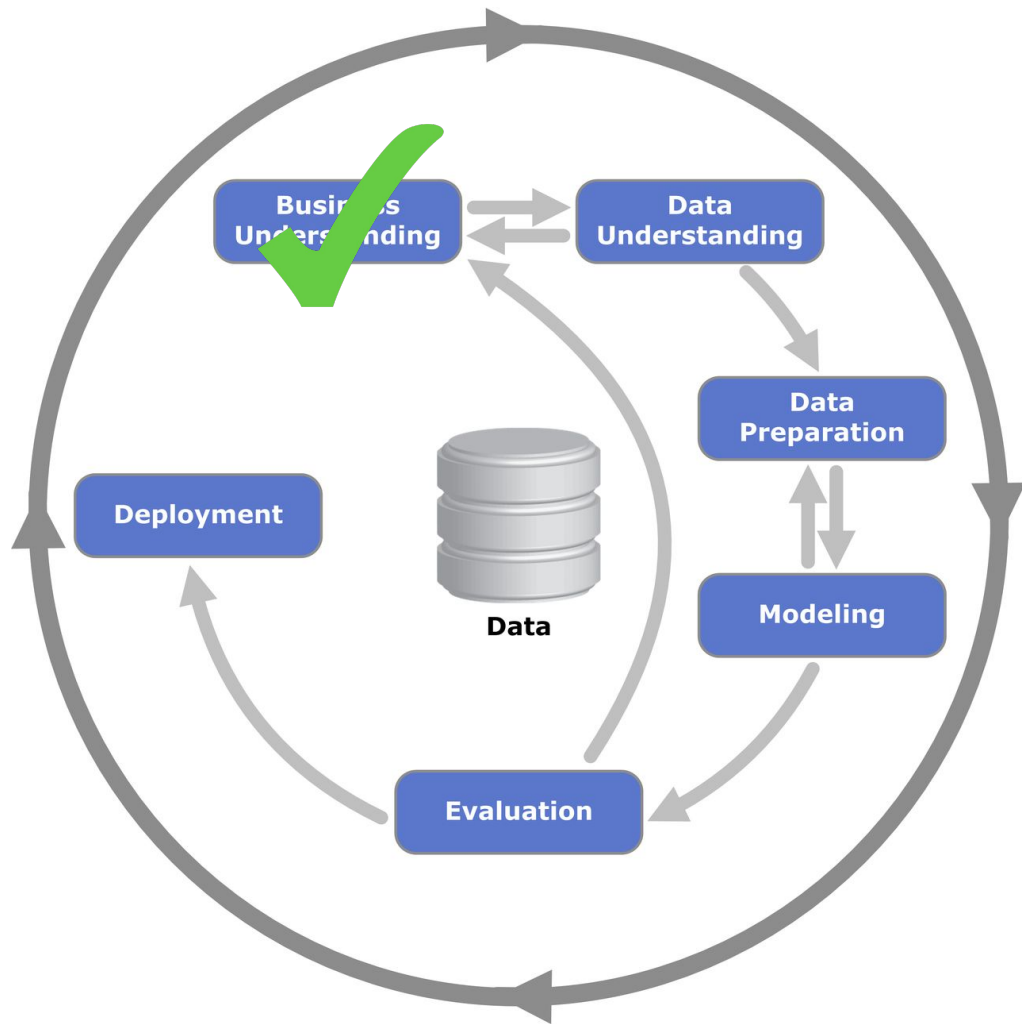
Frete grátis

COMPRA INTERNACIONAL

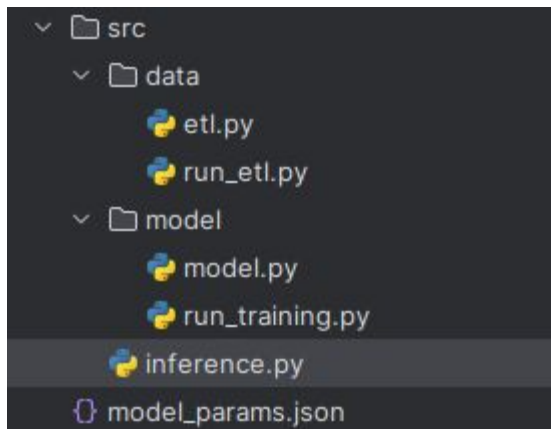
Vendedor local
x
Vendedor estrangeiro

Fluxo de Detecção





Implementação



Novo | 2235 vendidos

Smart TV LG AI ThinQ
60UP7750PSB LCD 4K 60"
100V/240V

★★★★★ (99)

 **RECOMENDADO** em Televisores

R\$ 5.599
R\$ 3.399 39% OFF
em 10x R\$ 339⁹⁰ sem juros
[Ver os meios de pagamento](#)

25% vendido Restam 05:07:12



Características gerais

Marca	LG
Linha	AI ThinQ
Modelo	60UP7750
Modelo alfanumérico	60UP7750PSB
Voltagem	100V/240V
Frequência	50 Hz x 60 Hz

Especificações

É smart	Sim
Quantidade de núcleos do processador	4
Sistema operativo	webOS 6.0
Com desligamento automático	Sim
Com sistema de montagem VESA incorporado	Sim

Peso e dimensões

Largura x Profundidade x Altura	1356 mm x 575 mm x 783 mm
---------------------------------	---------------------------

Tela

Tipo de tela	LCD
Tamanho da tela	60 "
Tipo de resolução	4K
Com HDR	Sim
Resolução máxima	3840 px - 2160 px
Relação de aspecto	16:9
Taxa de atualização da tela	60 Hz

Conectividade

Com USB	Sim
Com HDMI	Sim
Quantidade de portas HDMI	3
Quantidade de portas USB	2
Com Wi-Fi	Sim
Com Bluetooth	Sim
Com ethernet	Sim

Descrição

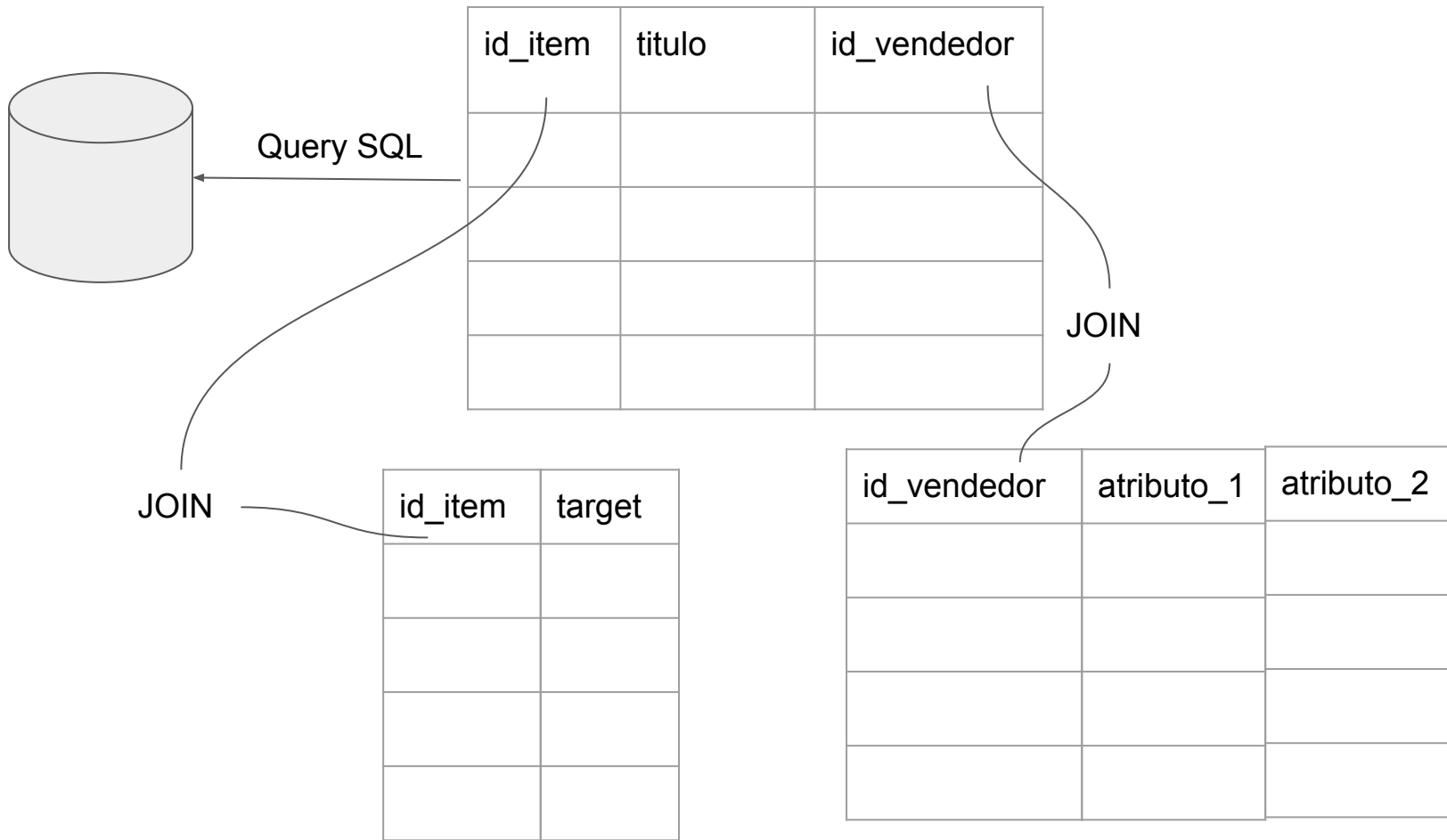
A LG é inovação e isso pode ser visto em cada um dos seus produtos tecnológicos, especialmente pensados para que toda a família aproveite mais a vida. Ter uma televisão LG é tirar proveito da mais alta qualidade do mercado. Com Smart TV 60UP7750 você poderá acessar os aplicativos onde seu conteúdo favorito está localizado. Além disso, você pode navegar na web, interagir em redes sociais e se divertir com jogos de vídeo.

Eu vivi em 4K

O número de pixels que oferece é 4 vezes maior do que Full HD, o resultado? Cenas muito mais realistas com um nível incrível de detalhes. Agora você vai conhecer uma aventura imersiva que não deixará de surpreendê-lo.

ETL (Extract, Transform and Load)

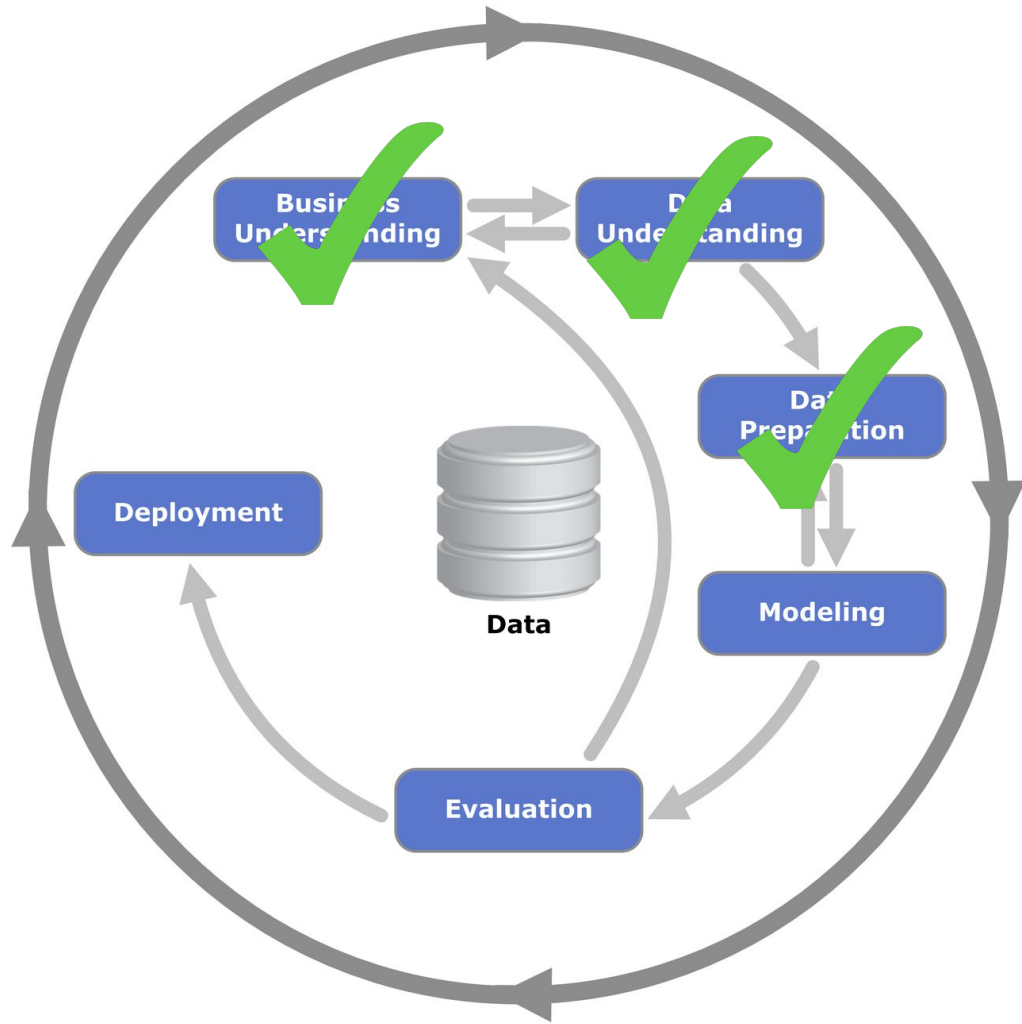
- Extract: Extração de dados de múltiplas fontes
 - Consulta a tabelas em data lakes e bases relacionais (SQL) ou arquivos (parquet, csv)
- Transform: Transformação dos dados para adequação
 - Criação de atributos/variáveis a serem usadas na criação de um modelo
- Load: Carga dos dados para a plataforma
 - Criação de um artefato de dados que possa ser carregado em memória



	site_id	item_title	item_context-domain_id	item_context-price_usd	seller_context-age_days	seller_feature_0	seller_feature_1	seller_feature_2
31269	MLM	Funda Elástica Para Sofá De 1/2/3/4 Plazas Tre...	MLM-SOFA_AND_FUTON_COVERS	32.053505	3.0	0.004985	0.011423	-0.010536
32543	MLM	Medias Altas Del Muslo Del Muslo De La Costill...	MLM-PANTYHOSE	24.149137	549.0	0.007624	0.007237	-0.012137
207870	MCO	6-18 Pulgadas Ropa De Muñeca Trajes Para 43 Cm...	MCO-BABY_ACTIVITY_TABLES	17.927133	0.0	0.004428	0.005762	-0.017597
62522	MLM	Esmalte De Uñas H Diamond Cat Eye Gel Esmalte ...	MLM-NAIL_STRENGTHENERS	21.062494	420.0	0.020709	0.011407	-0.007231

ETL

- Atributos:
 - Dataset name
 - Version
 - Selected features
- Métodos:
 - load_dataset_from_file
 - validate_dataset
 - serialize_dataset
 - deserialize_dataset



Treinamento

- Investigação e escolha de um algoritmo de aprendizado de máquina
- Treinamento do modelo utilizando os dados preparados
- Avaliação do modelo
- Criação de um artefato com o objeto contendo o modelo

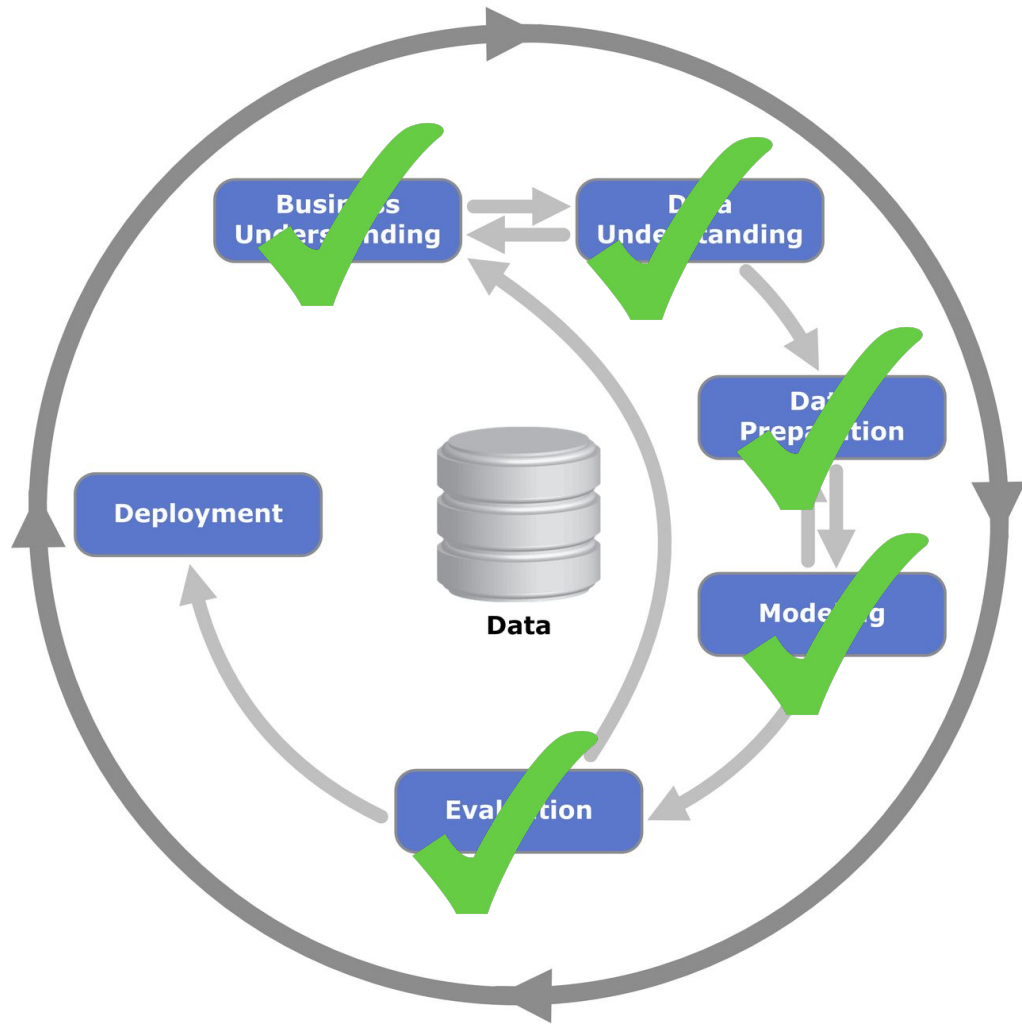
Model

- Atributos:

- name
- version
- model (objeto)
- metrics
- example
- *text_vectorizer*
- *text_feature_position*

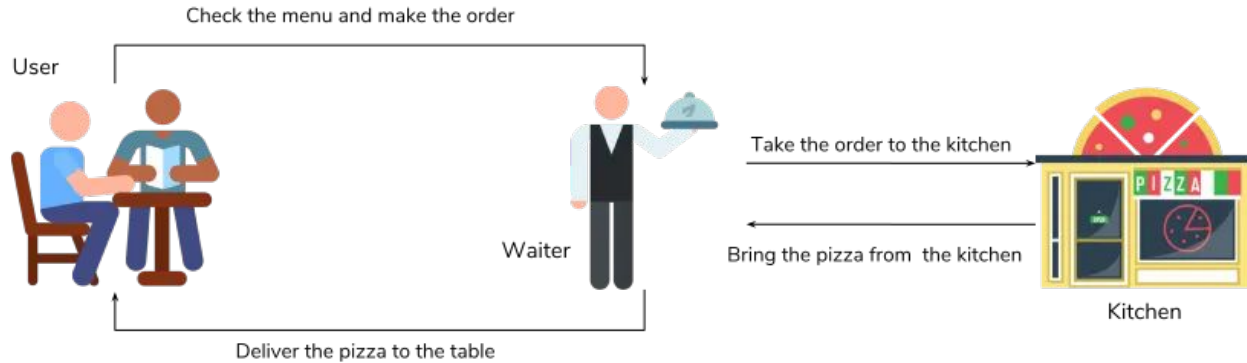
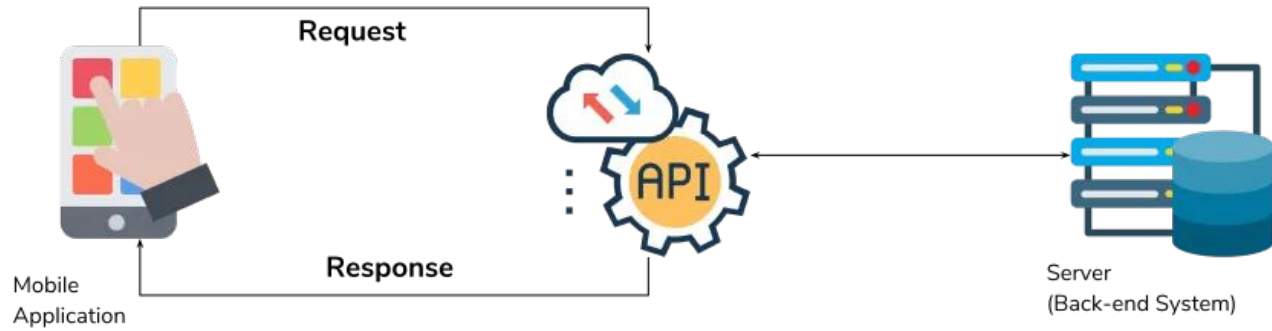
- Métodos:

- create_train_test_split
- train
- predict
- evaluate
- validate
- serialize
- load
- store_example
- *preprocess_text*

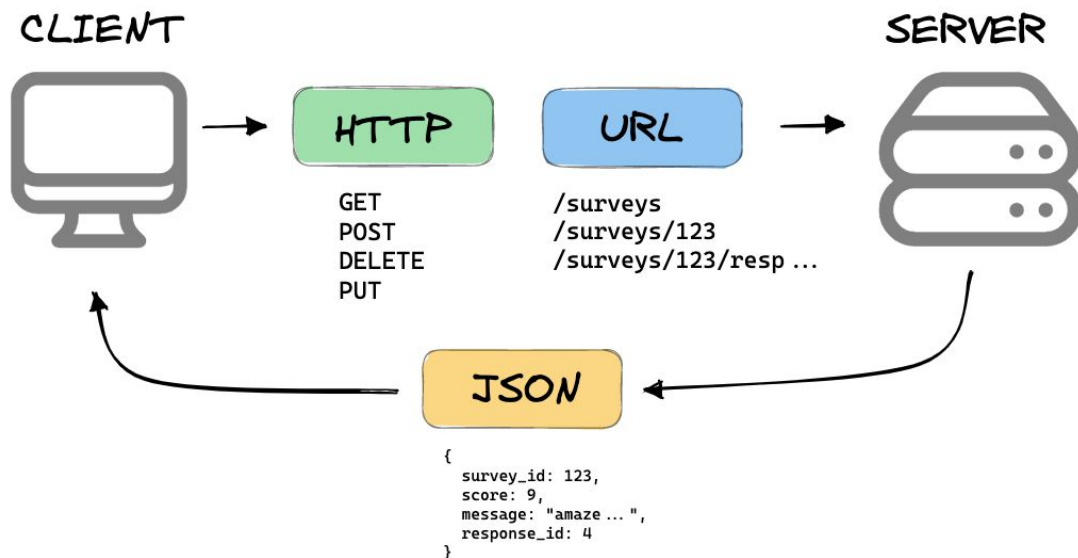


Inferência

- Código responsável por receber requisições e gerar respostas
- Contrato deve ser estabelecido de forma antecipada para integrar à plataforma
- Geralmente disponibilizado por um endpoint seguindo REST
- Deve ser eficiente



WHAT IS A REST API?

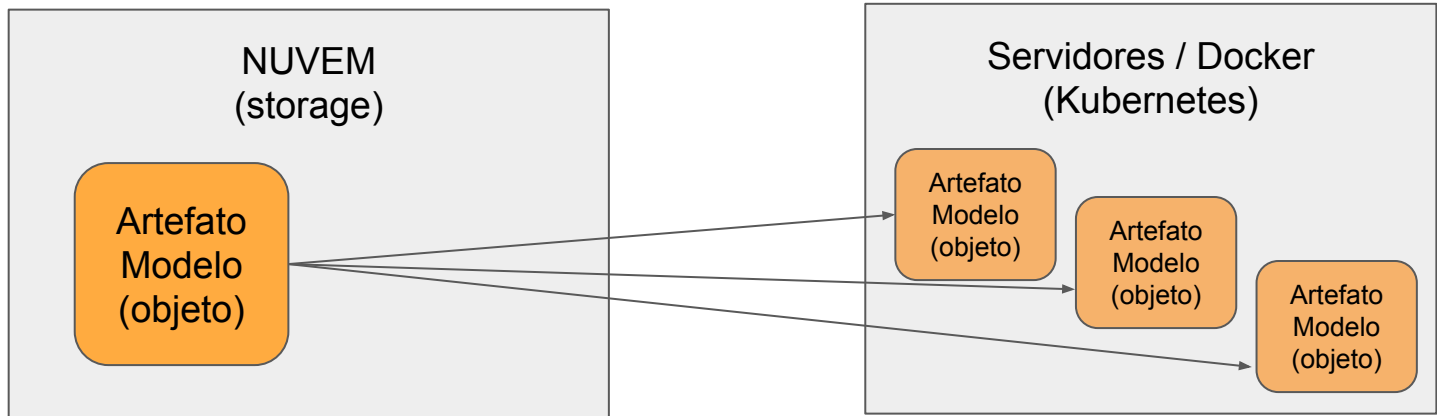


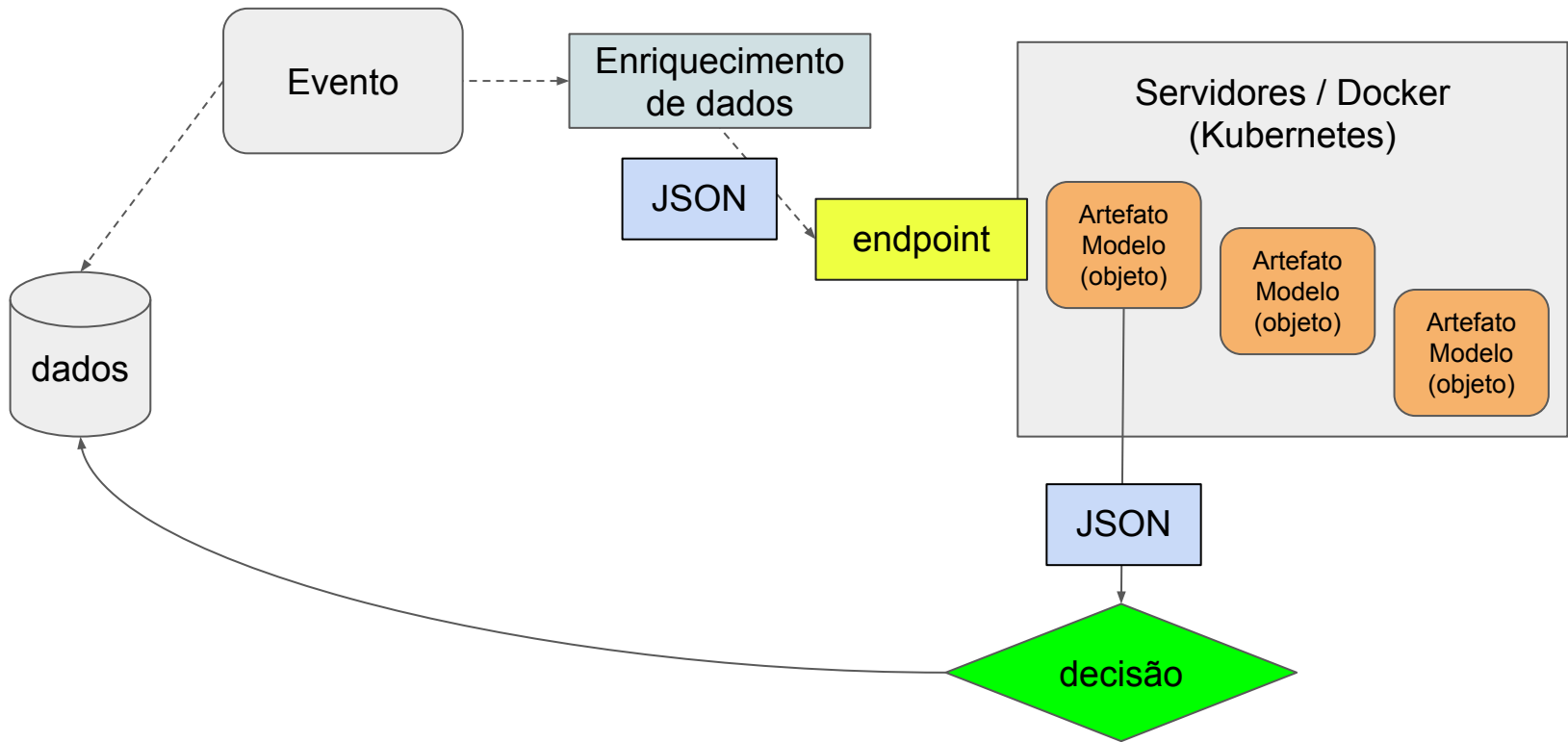
Status code	Meaning
200 OK	Request was successful.
301 Moved Permanently	For SEO purposes when a page has been moved and all link equity should be passed through.
401 Unauthorized	Server requires authentication.
403 Forbidden	Client authenticated but does not have permissions to view resource.
404 Not Found	Page not found because no search results or may be out of stock.
500 Internal Server Error	Server side error. Usually due to bugs and exceptions thrown on the server side code.
503 Server Unavailable	Server side error. Usually due to a platform hosting, overload and maintenance issue.

NUVEM
(storage)

Artefato
Modelo
(objeto)

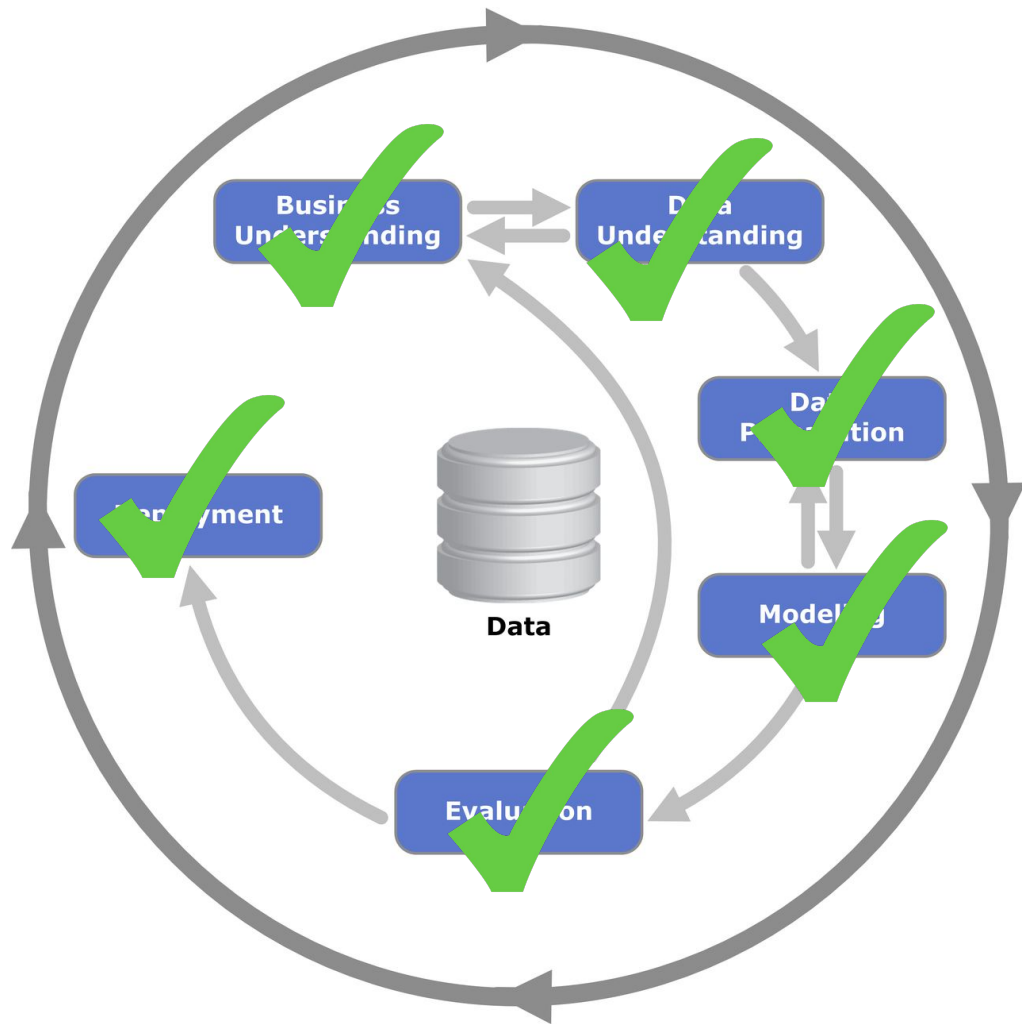
Servidores / Docker
(Kubernetes)





Inference

- Funções:
 - load_model
 - predict





<https://github.com/maponti>



[moacir-antonelli-ponti](#)



<https://sites.google.com/site/moacirponti/>