



Analytics e Inteligência Artificial

Aula 20

Deploy de Modelos





BUSINESS SCHOOL

Graduação, pós-graduação, MBA, Pós- MBA, Mestrado Profissional, Curso In Company e EAD



CONSULTING

Consultoria personalizada que oferece soluções baseadas em seu problema de negócio



RESEARCH

Atualização dos conhecimentos e do material didático oferecidos nas atividades de ensino



Líder em Educação Executiva, referência de ensino nos cursos de graduação, pós-graduação e MBA, tendo excelência nos programas de educação. Uma das principais escolas de negócio do mundo, possuindo convênios internacionais com Universidades nos EUA, Europa e Ásia. +8.000 projetos de consultorias em organizações públicas e privadas.



Único curso de graduação em administração a receber as notas máximas



A primeira escola brasileira a ser finalista da maior competição de MBA do mundo



Única Business School brasileira a figurar no ranking LATAM



Signatária do Pacto Global da ONU



Membro fundador da ANAMBA - Associação Nacional MBAs



Credenciada pela AMBA - Association of MBAs



Credenciada ao Executive MBA Council



Filiada a AACSB - Association to Advance Collegiate Schools of Business



Filiada a EFMD - European Foundation for Management Development



Referência em cursos de MBA nas principais mídias de circulação

O Laboratório de Análise de Dados - LABDATA é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de *Big Data*, *Analytics* e *Inteligência Artificial*.



Profª Drª Alessandra Montini

O LABDATA é um dos pioneiros no lançamento dos cursos de *Big Data* e *Analytics* no Brasil

Os diretores foram professores de grandes especialistas do mercado

+10 anos de atuação

+1000 alunos formados

Docentes

- Sólida formação acadêmica: doutores e mestres em sua maioria
- Larga experiência de mercado na resolução de *cases*
- Participação em Congressos Nacionais e Internacionais
- Professor assistente que acompanha o aluno durante todo o curso

Estrutura

- 100% das aulas realizadas em laboratórios
- Computadores para uso individual durante as aulas
- 5 laboratórios de alta qualidade (investimento +R\$2MM)
- 2 Unidades próximas a estação de metrô (com estacionamento)



Corpo Diretivo

COORDENADORES DO LABDATA | ATUAÇÃO ACADÊMICA E PROFISSIONAL

4



Profª Dra.
Alessandra Montini

Diretora do LABDATA-FIA, apaixonada por dados e pela arte de lecionar. Têm muito orgulho de ter criado na FIA cinco laboratórios para as aulas de Big Data e inteligência Artificial. Possui mais de 20 anos de trajetória nas áreas de Data Mining, Big Data, Inteligência Artificial e Analytics. Cientista de dados com carreira realizada na Universidade de São Paulo. Graduada e mestra em estatística aplicada pelo IME-USP e doutora pela FEA-USP. Com muita dedicação chegou ao cargo de professora e pesquisadora na FEA-USP, ganhou mais de 30 prêmios de excelência acadêmica pela FEA-USP e mais de 30 prêmios de excelência acadêmica como professora dos cursos de MBA da FIA. Orienta alunos de mestrado e de doutorado na FEA-USP. Membro do Conselho Curador da FIA, Coordenadora de Grupos de Pesquisa no CNPQ, Parecerista da FAPESP e Colunista de grandes Portais de Tecnologia.

 [linkedin.com/in/alessandramontini/](https://www.linkedin.com/in/alessandramontini/)



Prof. Dr.
Adolpho Walter Canton

Diretor do LABDATA-FIA. Consultor em Projetos de *Analytics*, *Big Data* e Inteligência Artificial. Professor FEA - USP. PhD em Estatística Aplicada pela *University of North Carolina at Chapel Hill*, Estados Unidos.



Currículo - Prof. João Nogueira

FORMAÇÃO ACADÊMICA | EXPERIÊNCIA PROFISSIONAL

- (2019-Presente) - Professor nos cursos de Extensão, Pós e MBA em Big Data e Data Mining na Fundação Instituto de Administração (FIA) - www.fia.com.br
- (2018-Presente) - Cientista de Dados na Via Varejo - <https://viavarejo.com.br>
- (2016-Presente) - Doutorando em Física Computacional e Estatística pelo Departamento de Física na Universidade Federal do Ceará - <https://fisica.ufc.br>
- (2014-2016) - Mestre em Física da Matéria Condensada pelo Departamento de Física na Universidade Federal do Ceará - <https://fisica.ufc.br>
- (2012-2013) - Estudante Intercambista na Universidade de Coimbra - Portugal - <https://www.uc.pt>
- (2010-2014) - Bacharel em Física pela Universidade Federal do Ceará - <http://www.ufc.br>
- Contatos:
 - E-mail: joaonogueira@fisica.ufc.br



Currículo - Prof. Dino Magri

FORMAÇÃO ACADÊMICA | EXPERIÊNCIA PROFISSIONAL

6

- **(2014-Presente)** - Professor nos cursos de Extensão, Pós e MBA em Big Data e Data Mining na Fundação Instituto de Administração (FIA) - www.fia.com.br
- **(2018-Presente)** - Pesquisa e Desenvolvimento de Big Data e Machine Learning na Beholder (<http://beholder.tech>)
- **(2013-2018)** - Pesquisa e Desenvolvimento no Laboratório de Arquitetura e Redes de Computadores (LARC) na Universidade de São Paulo - www.larc.usp.br
- **(2012)** - Bacharel em Ciência da Computação pela Universidade do Estado de Santa Catarina (UDESC) - www.cct.udesc.br
- **(2009/2010)** - Pesquisador e Desenvolvedor no Centro de Computação Gráfica - Guimarães - Portugal - www.ccg.pt
- **Contatos:**
 - E-mail: professor.dinomagri@gmail.com
 - Site: <http://www.dinomagri.com>
 - Twitter: <https://twitter.com/dinomagri>
 - Lattes: <http://lattes.cnpq.br/5673884504184733>



Conteúdo Programático da Disciplina - Projeto de Inteligência Artificial



| Data | Horário | Tema |
|------------|---------|---|
| 09/03/2021 | 19:00 | Aula 1 - Introdução ao Ambiente de Desenvolvimento |
| 11/03/2021 | 19:00 | Aula 2 - Revisão de Python |
| 16/03/2021 | 19:00 | Aula 3 - Manipulação de Dados |
| 18/03/2021 | 19:00 | Aula 4 - Análise Exploratória de Dados |
| 23/03/2021 | 19:00 | Aula 5 - Projeto da disciplina - Parte 1 - Análise Exploratória de Dados |
| 25/03/2021 | 19:00 | Aula 6 - Introdução, Motivação e Framework de Machine Learning |
| 06/04/2021 | 19:00 | Aula 7 - Analytical Base Table |
| 08/04/2021 | 19:00 | Aula 8 - Aprendizagem Supervisionada - Classificação |
| 13/04/2021 | 19:00 | Aula 9 - Aprendizagem Supervisionada - Classificação |
| 15/04/2021 | 19:00 | Aula 10 - Aprendizagem Supervisionada - Classificação |
| 20/04/2021 | 19:00 | Aula 11 - Projeto da disciplina - Parte 2 - Machine Learning - Classificação |
| 22/04/2021 | 19:00 | Aula 12 - Projeto da disciplina - Parte 2 - Machine Learning - Classificação |
| 27/04/2021 | 19:00 | Aula 13 - Aprendizagem Supervisionada - Regressão |
| 29/04/2021 | 19:00 | Aula 14 - Aprendizagem Supervisionada - Regressão |
| 04/05/2021 | 19:00 | Aula 15 - Projeto da disciplina - Parte 3 - Machine Learning - Regressão |
| 06/05/2021 | 19:00 | Aula 16 - Aprendizagem Não-Supervisionada |
| 11/05/2021 | 19:00 | Aula 17 - Aprendizagem Não-Supervisionada |
| 13/05/2021 | 19:00 | Aula 18 - Projeto da disciplina - Parte 4 - Machine Learning - Clusterização |
| 18/05/2021 | 19:00 | Aula 19 - AutoML |
| 20/05/2021 | 19:00 | Aula 20 - Demonstração de Deploy de Machine Learning |

Conteúdo da Aula

- 1. Introdução
- 2. Deploy do Modelo



Material das aulas

- Iremos utilizar o Google Colab para desenvolver os códigos durante as aulas.
- Acesse <https://bit.ly/tutorial-colab-projeto> para realizar o tutorial de utilização do Google Colab.



1. Introdução



1. Introdução

INTRODUÇÃO

O que significa implementar (deploy) um modelo de Machine Learning?



1. Introdução

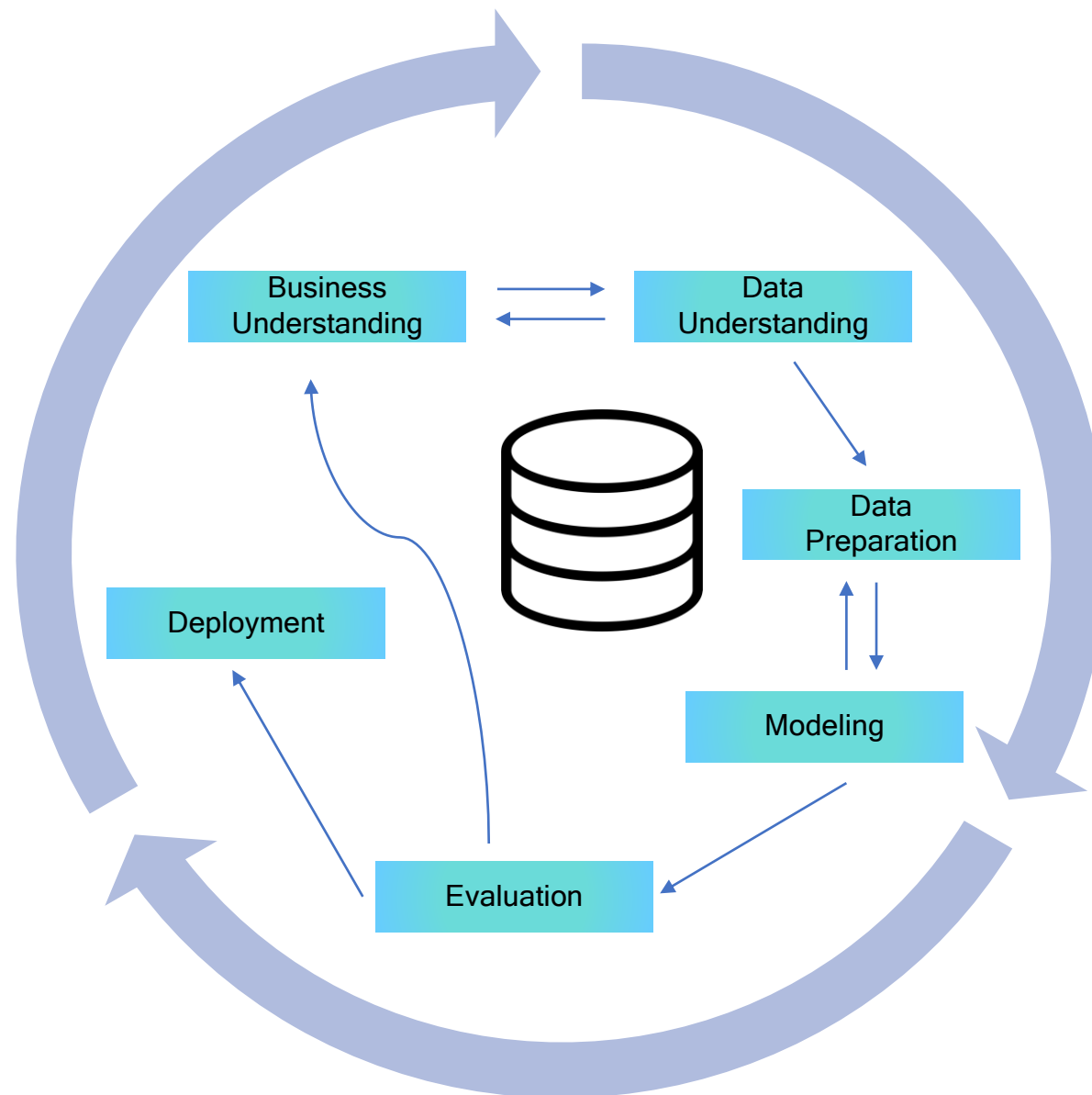
INTRODUÇÃO

- Significa envolver diversas equipes nesse processo



1. Ciclo de vida de um projeto de Machine Learning

INTRODUÇÃO



1. Introdução

INTRODUÇÃO

- Um projeto de Machine Learning deve existir para criar valor dentro de uma empresa.
- O processo de treinar um modelo de Machine Learning e disponibilizar as predições para os usuários ou outros sistemas, é conhecido como implementação/deploy.
- Como você decide implementar um modelo de Machine Learning?



1. Introdução

INTRODUÇÃO

- Para decidir é necessário entender como os usuários finais devem interagir com as previsões realizadas pelo modelo. Isso significa ter um caso de uso bem definido.
- Isso irá impactar no deploy dos modelos, uma vez que envolve múltiplas equipes, arquiteturas e ferramentas.
- É muito importante que o ponto de partida na hora de construir uma arquitetura seja sempre levar em considerações os requisitos de negócio e **objetivos mais amplos da empresa**.
- As restrições também são importantes, bem como qual o valor que está criando e para quem.



1. Introdução

INTRODUÇÃO

- Algumas perguntas que podem auxiliar nesse processo inicial:
 - É necessário fornecer previsões em tempo real? Se sim, qual a latência da aplicação (milissegundos, segundos, minutos, dias, etc)?
 - É necessário retreinar sempre o modelo? Se sim, qual a periodicidade do retreino?
 - Qual o tamanho dos dados que serão utilizados na predição? Será uma predição unitária ou em várias amostras de uma vez (batch)?
 - Qual o tipo de algoritmo será utilizado? Algoritmos como regressão linear, logística e k-means muitas vezes podem ser implementados diretamente no local onde os dados estão armazenados.
 - Quantas aplicações e usuários irão acessar os modelos?
 - Qual a latência mínima dessas aplicações?
 - Qual o tamanho e experiência de sua equipe?
 - Quais outras perguntas?



1. Introdução

INTRODUÇÃO

- Com os requisitos, é possível pensar em algumas arquiteturas de alto nível para sistemas de ML.

| | REST API | Banco de dados | Streaming | Aplicação móvel (Apps) |
|--|--------------|--------------------|--------------------------------|-------------------------------|
| Treinamento | Batch | Batch | Streaming | Streaming |
| Predição | Imediata | Batch | Streaming | Imediata |
| Entrega do resultado da predição | Via REST API | Via Banco de dados | Streaming via fila de mensagem | Via API processamento celular |
| Latência da predição | OK | Alta | Muito baixa | Baixa |
| Dificuldade de gerenciamento de sistemas | OK | Simples | Muito Alta | OK |



1. Introdução

INTRODUÇÃO

- **Princípios** chaves para projetar um sistema de Machine Learning
 - **Reprodutibilidade desde o início**
 - Persista todas as entradas e saídas do modelos, bem como os *metadados* (configurações, dependências, etc.)
 - Utilize um sistema de versionamento, tanto do código, quanto dos dados
 - **Trate as etapas de ML como parte da compilação**
 - Automatize o treinamento e a publicação dos modelos
 - **Planeje a extensibilidade**
 - Atualização dos modelos deve ser realizado com cautela principalmente se as atualizações forem regulares



1. Introdução

INTRODUÇÃO

- **Princípios** chave para projetar um sistema de Machine Learning
 - **Modularidade**
 - Se possível, reutilize o código de pré-processamento e da seleção de atributos utilizados no ambiente de desenvolvimento para o ambiente de produção
 - **Teste**
 - É uma parte muito importante quando se pretende automatizar qualquer sistema. Eles são cruciais.
 - Testes diferenciais
 - Testes de Benchmark
 - Testes de carga e estresse



1. Introdução

INTRODUÇÃO

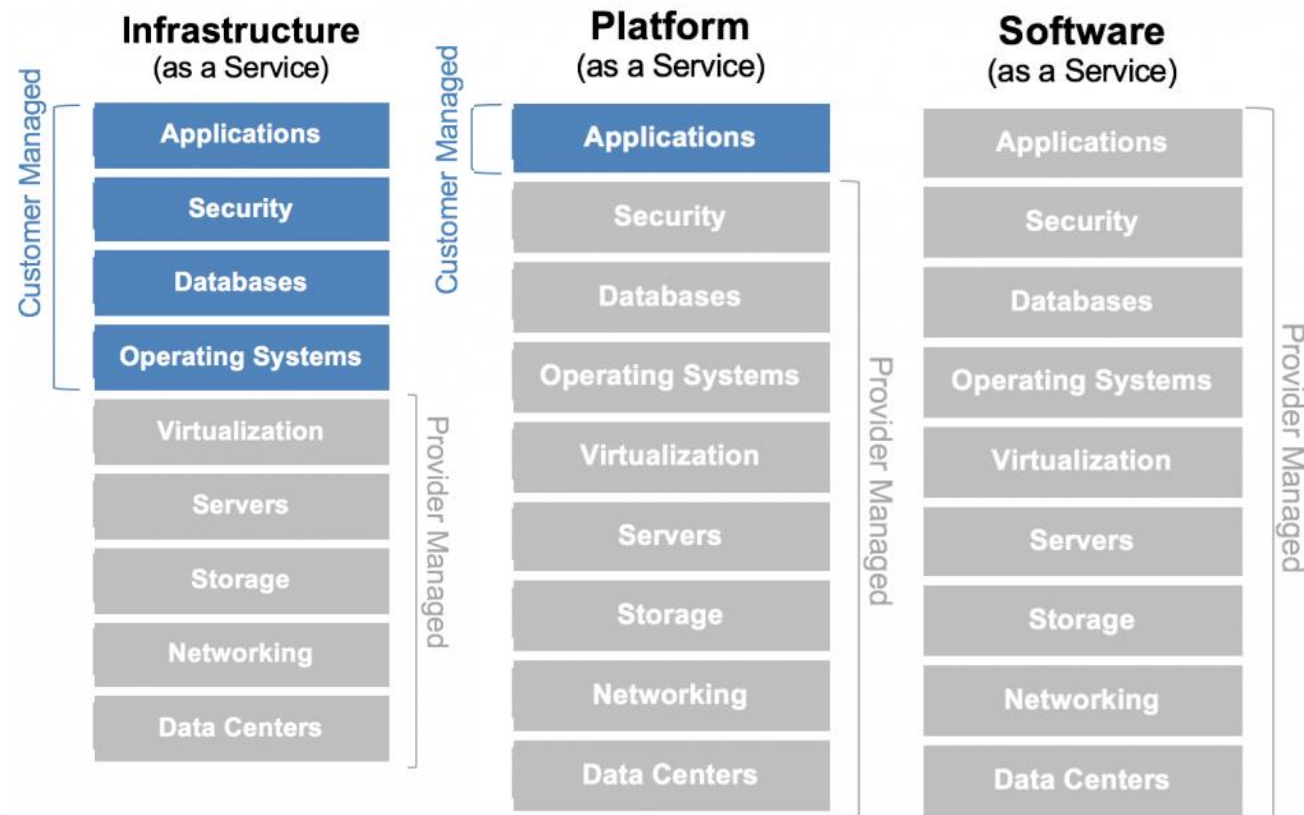
- Em relação aos testes, é muito importante e vale citar alguns testes relacionados aos modelos de ML
 - **Testes diferenciais**: comparação da performance do novo modelo (desafiante) com a performance dos modelos antigos em um mesmo conjunto de dados de teste.
 - **Testes de Benchmark**: Tempo gasto no treinamento e nas predições em comparação ao modelo anterior.
 - **Testes de carga e estresse**: Não são específicos de ML, porém vale a pena executar esse tipo de teste uma vez que existem altas demandas por uso de CPU/GPU e Memória para aplicações de ML.
- Outros testes podem ser visualizados em <https://landing.google.com/sre/sre-book/chapters/testing-reliability/>



1. Introdução

INTRODUÇÃO

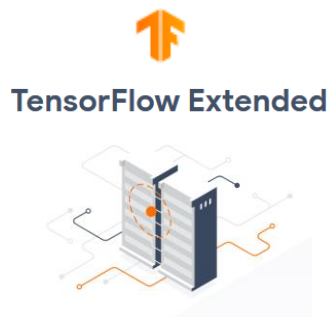
- Com isso, já podemos pensar em que tipo de infraestrutura a implantação será realizada:
 - Plataforma como serviço (PaaS)
 - Infraestrutura como serviço (IaaS)



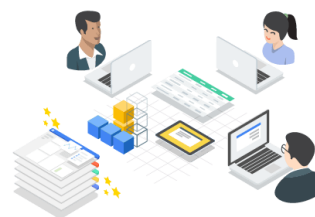
1. Introdução

INTRODUÇÃO

- Existem diversas soluções no mercado que mantem serviços específicos para implementação de modelos de Machine Learning.



AI Platform



Azure Machine Learning



1. Introdução

INTRODUÇÃO

- Esses serviços fornecem um pipeline de processamento dos dados que permite reproduzir uma execução em diversos ambientes.
- Isso facilita, por exemplo, o teste dentro de um ambiente de desenvolvimento, homologação e produção.
- Dentro desses pipelines:
 - Acessando as fontes de dados
 - Pré-processamento dos dados
 - Selecionando variáveis
 - Construção do modelo
- Tanto o Spark como o Scikit-Learn permitem o uso de pipelines 😊



1. Introdução

INTRODUÇÃO

- O uso de pipelines permite uma automatização maior principalmente se atrelado as ferramentas de desenvolvimento e operacionalização de software existentes atualmente:
 - Containers
 - Fluxos de CI/CD (Continuous Integration/Continuous Development)
 - Estratégias de Implantação



1. Introdução

INTRODUÇÃO

- **Containers**

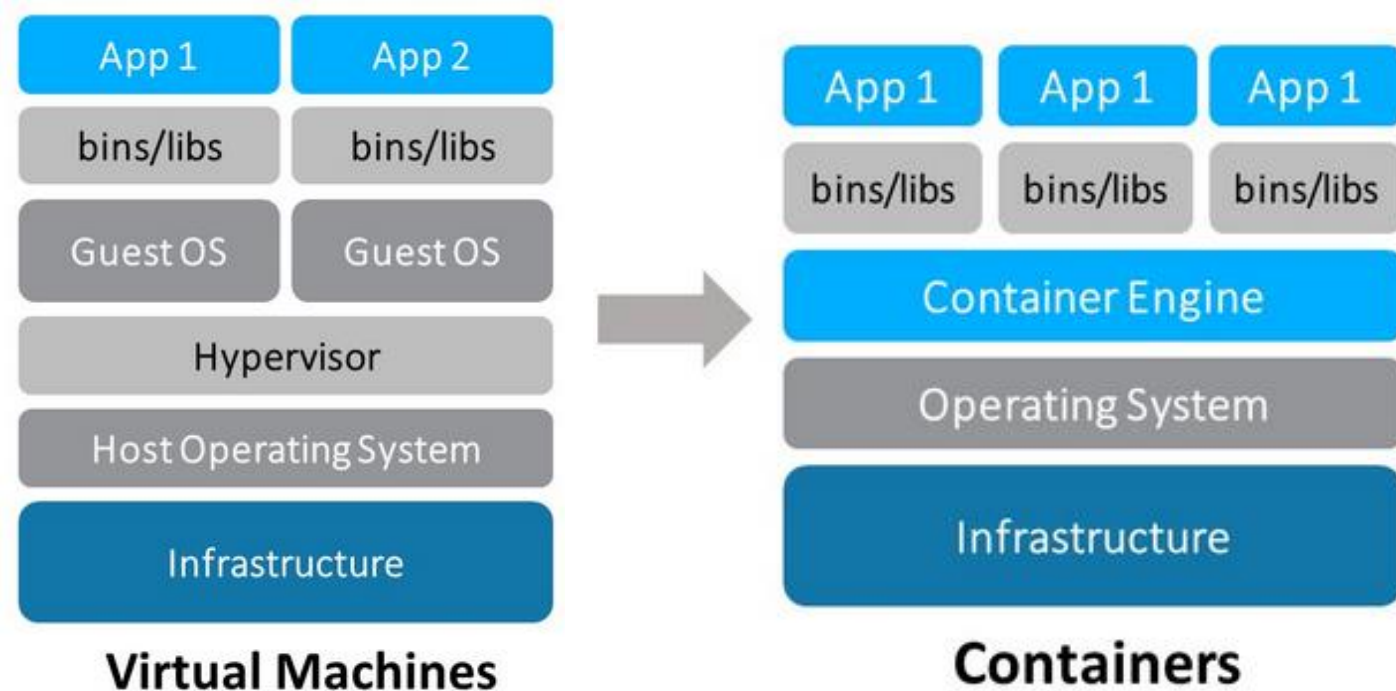
- Desde que surgiu em 2013, o Docker revolucionou a forma que um software é entregue em produção.
- O maior benefício de se criar uma aplicação usando contêineres é a garantia de que as dependências necessárias do projeto funcionem em qualquer sistema operacional.
- Isso permite criar ambientes de implementação precisos, sendo uma grande vantagem na hora de reproduzir, testar, treinar e portar os modelos de ML.
- Acaba com a velha máxima: “Na minha máquina funciona”.



1. Introdução

INTRODUÇÃO

- Containers



1. Introdução

INTRODUÇÃO

- **Fluxos de CI/CD**

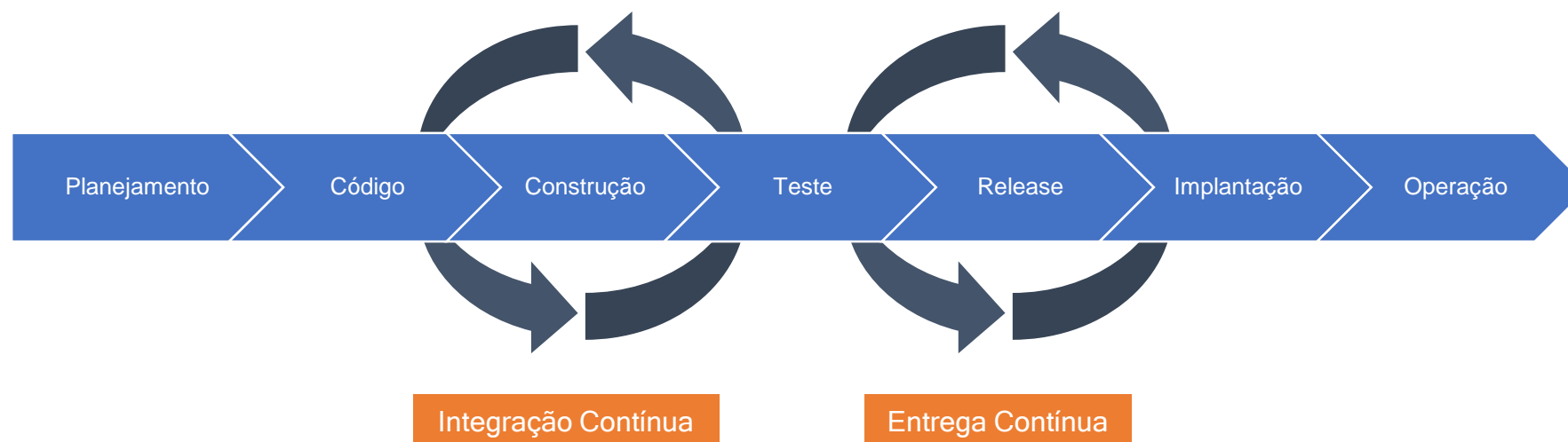
- CI/CD representa um conjunto de ferramentas e processos que facilitam a evolução do projeto de ML a partir do ambiente de desenvolvimento até o ambiente de produção, mitigando todos os riscos existentes nesse fluxo.
- Um modelo é quase que imediato colocado em produção, devido a automatização de todos os testes e procedimentos necessário.
- Exemplos de Ferramentas de CI/CD: Jenkins, Github Actions, CircleCI, Azure DevOps, Travis.



1. Introdução

INTRODUÇÃO

- Fluxos de CI/CD



1. Introdução

INTRODUÇÃO

- **Estratégias de Implementação**

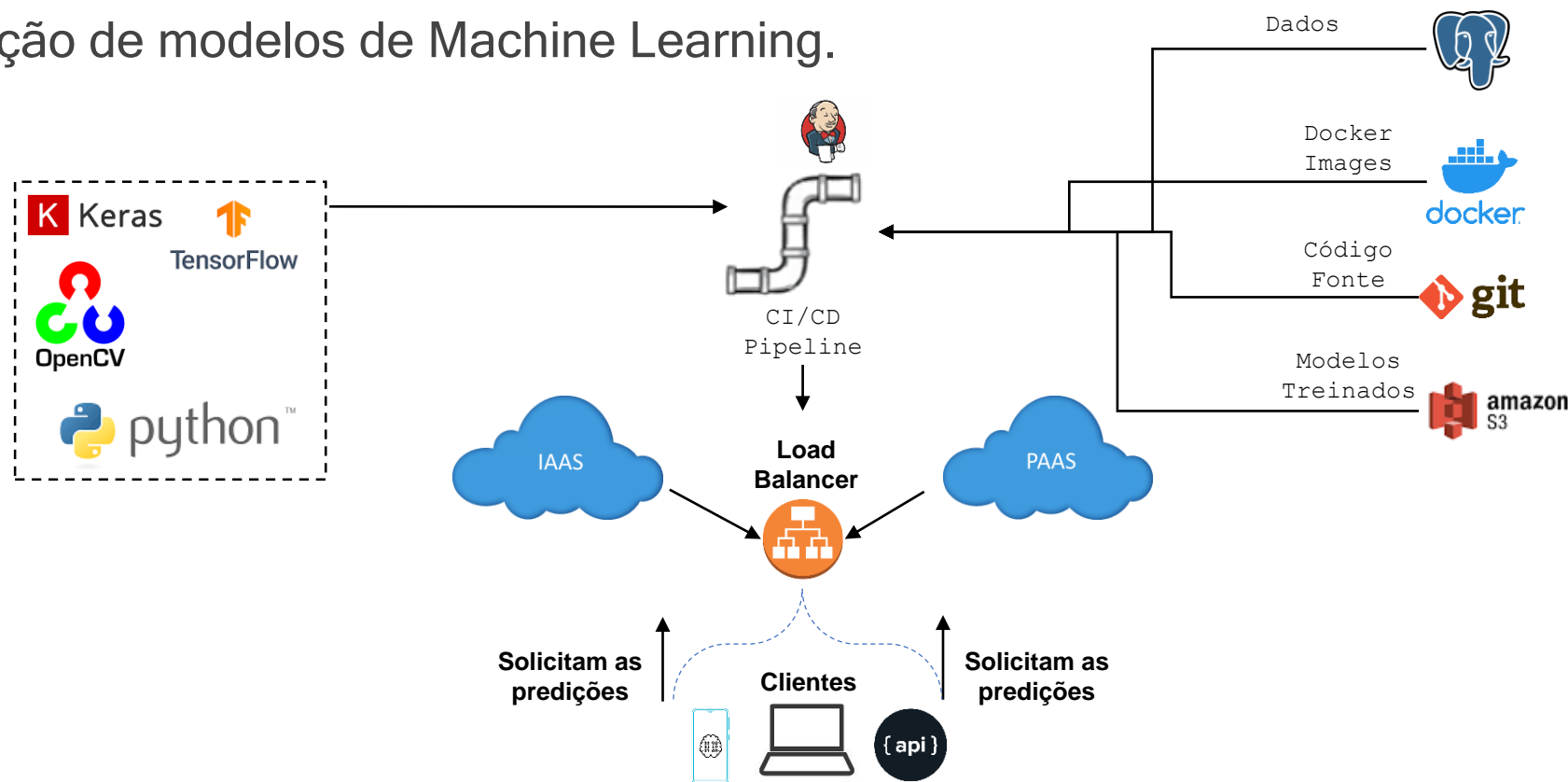
- **Shadow Mode** é uma forma de testarmos os modelos criados em produção, sem de fato ter atualizado o algoritmo antigo. A ideia é comparar as respostas obtidas pelo modelo antigo e novo em um ambiente de produção.
- **Canary Release** é uma forma de realizar a implantação dos modelos somente para uma parcela dos usuários do sistema de ML.
- **Post-Release** de fato é a disponibilização do novo modelo de ML para todos os usuários do sistema de ML.



1. Introdução

INTRODUÇÃO

- Como isso, podemos definir um fluxo simples que envolvem todos os componentes no ciclo de implantação de modelos de Machine Learning.



2. Deploy do Modelo



2. Deploy do Modelo

- O modelo deverá ser executado de forma automática todo dia 1 do mês
- As previsões serão disponibilizadas via API, dessa forma ficará disponível para outros usuários e sistemas (banco de dados) consumirem
- Também será disponibilizado um **web app** para que um usuário interaja de forma mais amigável com o modelo.
- Github do projeto: <https://github.com/joaopcnogueira/propensao-revenda>



2. Deploy do Modelo

CÓDIGO FONTE DO PROJETO

joaopcogueira / propensao-revenda

main 1 branch 0 tags

Go to file Add file Code

| File | Commit Message | Time |
|------------------|--------------------------|-------------|
| datasets | upload files | 3 days ago |
| docs | docs folder created | 3 days ago |
| models | upload files | 3 days ago |
| notebooks | update README | 3 days ago |
| .gitignore | upload files | 3 days ago |
| README.md | Update README.md | 6 hours ago |
| api.py | streamlit web app added | 3 days ago |
| deploy.sh | script de deploy | 3 days ago |
| requirements.txt | update requirements.txt | 3 days ago |
| web_app.py | File Download Adicionado | 5 hours ago |

- **datasets:** pasta onde ficam as bases de dados
- **docs:** pasta onde ficam guardadas quaisquer documentação gerada ou necessária para o processo de modelagem
- **models:** aqui ficam armazenados os modelos
- **notebooks:** pasta para os notebooks
- **.gitignore:** arquivo auxiliar que lista os arquivos ignorados pelo git
- **README.md:** documentação do projeto
- **api.py:** código da api
- **deploy.sh:** lista de comandos para deploy em um servidor
- **requirements.txt:** lista dos pacotes necessários para executar o projeto
- **web_app.py:** código do web app

2. Deploy do Modelo

FERRAMENTAS PARA EXECUTAR O PROJETO



Linguagem de programação



Gerenciamento de ambientes virtuais



Processamento de dados



Ambiente de treinamento



Versionamento de código



Processamento de dados



Construção de APIs



Repositório de código



Machine learning



Construção de Web Apps



Machine learning



2. Deploy do Modelo

API

FastAPI 0.1.0 OAS3

/openapi.json

default

GET / Hello

POST /predict/ Predict

Schemas

HTTPValidationError >

ValidationError >



2. Deploy do Modelo

WEB APP

×

Como você gostaria de fazer a sua predição?

Online

Previsão de Não Revenda

UF

SP

Total de Pedidos

5

Total de Items

5

Total de Items Distintos

5

Receita

5

Recência

5

Predict

A probabilidade do seller não revender nos próximos 6 meses é de 0.44

×

Como você gostaria de fazer a sua predição?

Batch

Previsão de Não Revenda

Upload csv file for predictions

Drag and drop file here

Limit: 200MB per file • CSV

Browse files



Referências Bibliográficas



Referências Bibliográficas

ARTIGOS

How to Deploy Machine Learning Models

A Guide

17 March 2019

Deploying Machine Learning Models in Shadow Mode

A Guide

30 March 2019

by: Soledad Galli | January 27, 2020

How To Build And Deploy A Reproducible Machine Learning Pipeline



by: Soledad Galli | February 15, 2020

How to Test and Monitor Machine Learning Model Deployments



Monitoring Machine Learning Models in Production

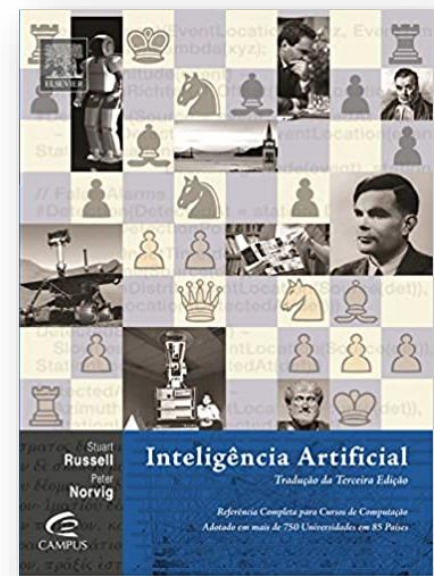
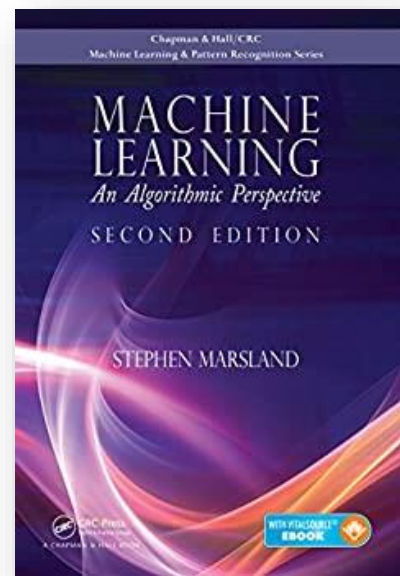
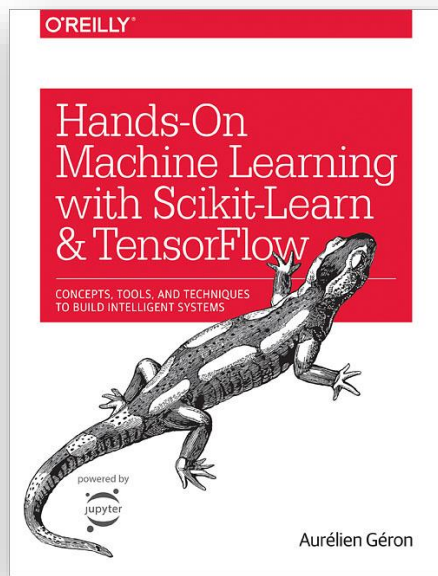
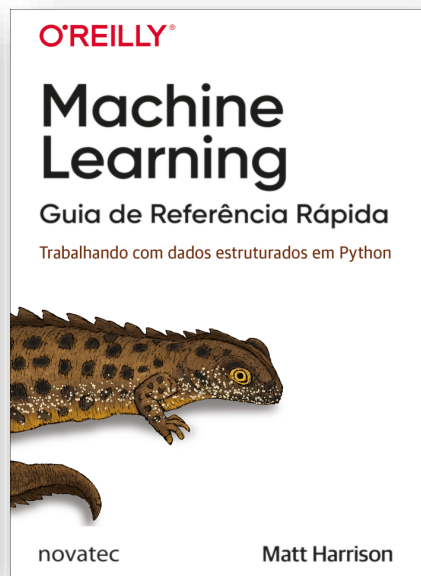
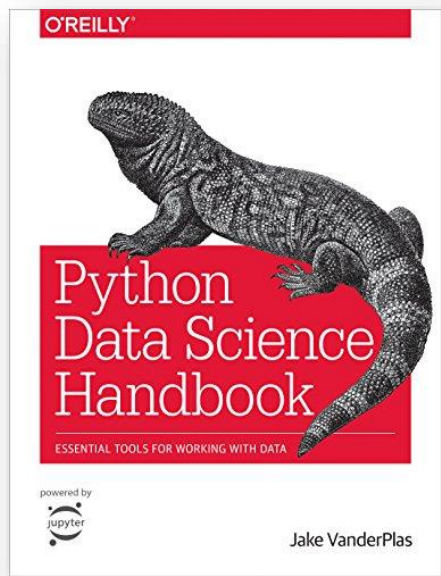
A Comprehensive Guide

14 March 2020



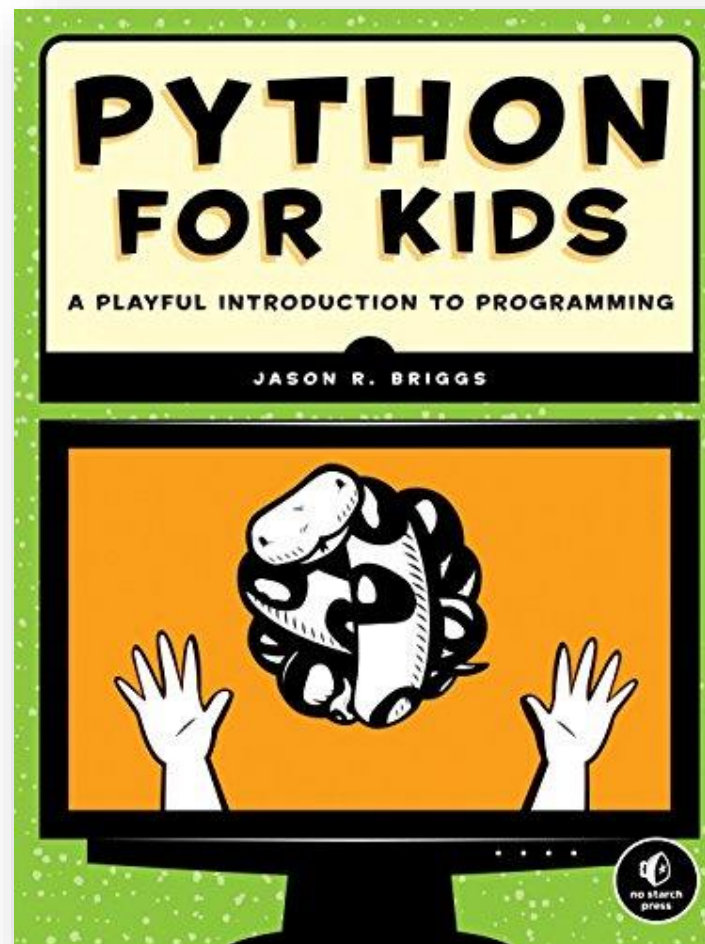
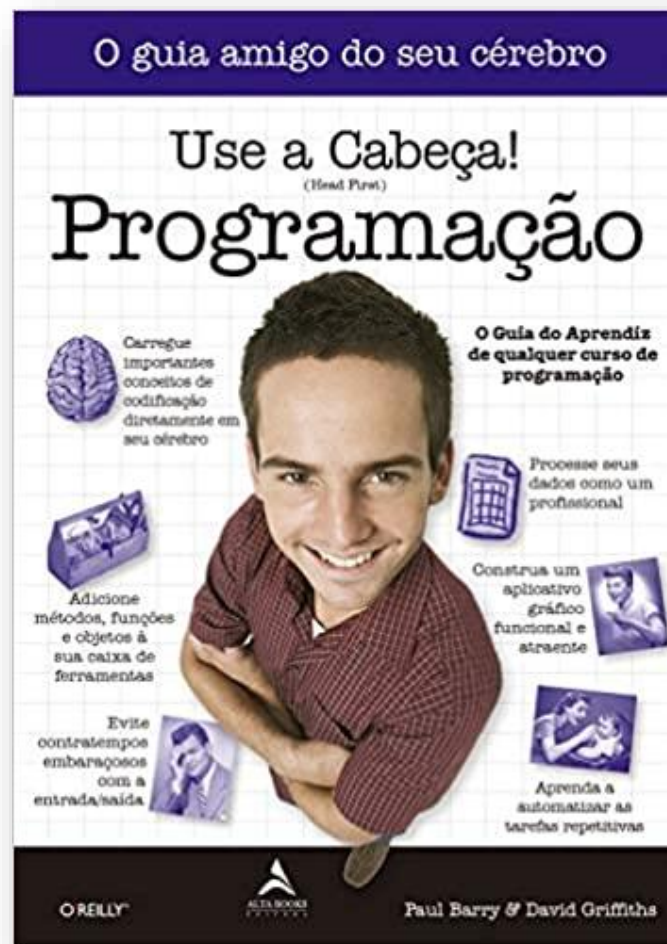
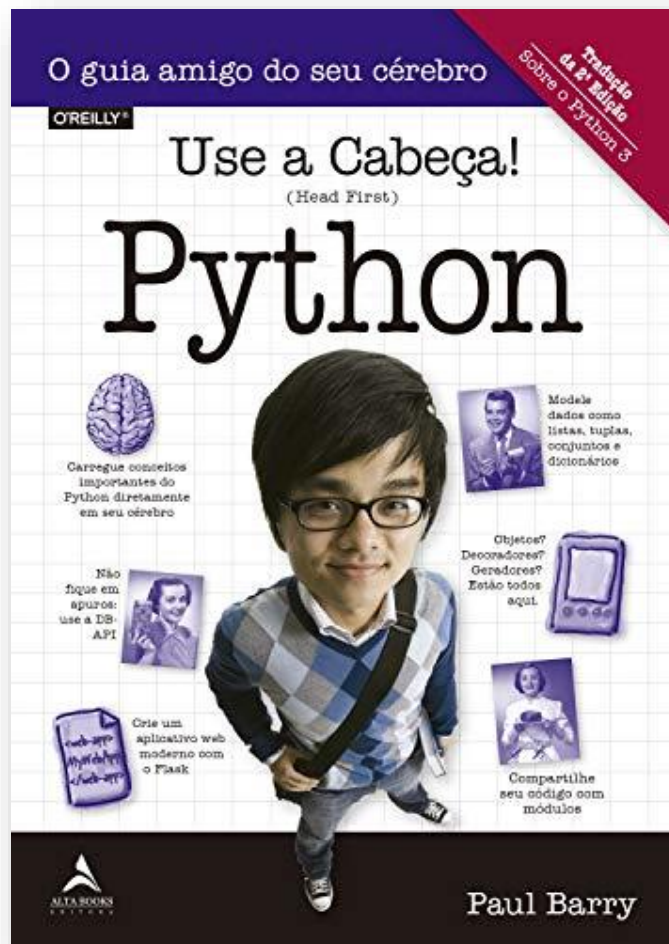
Referências Bibliográficas

LIVROS



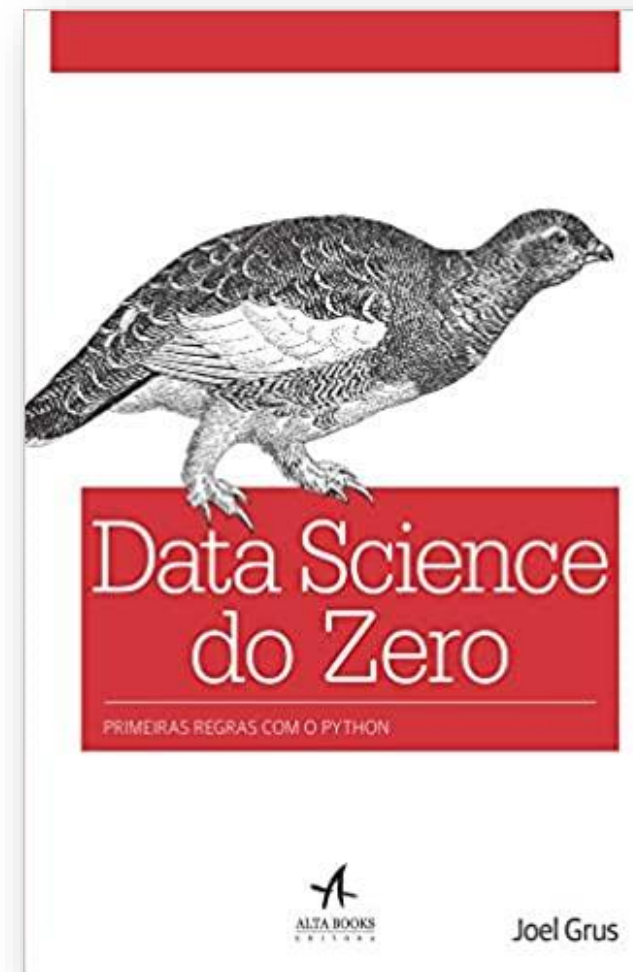
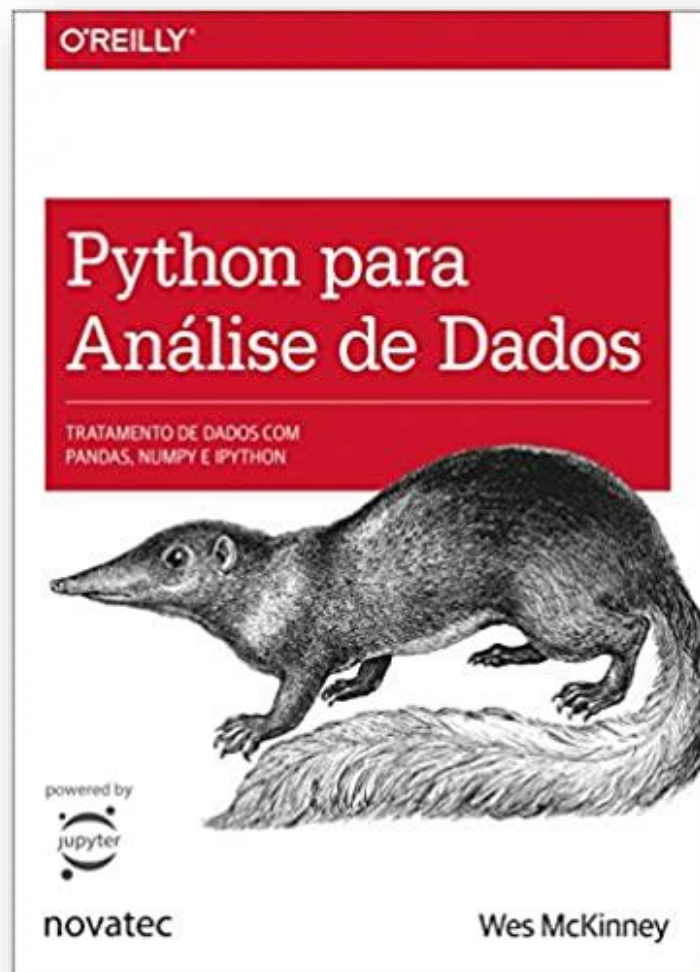
Referências Bibliográficas

LIVROS



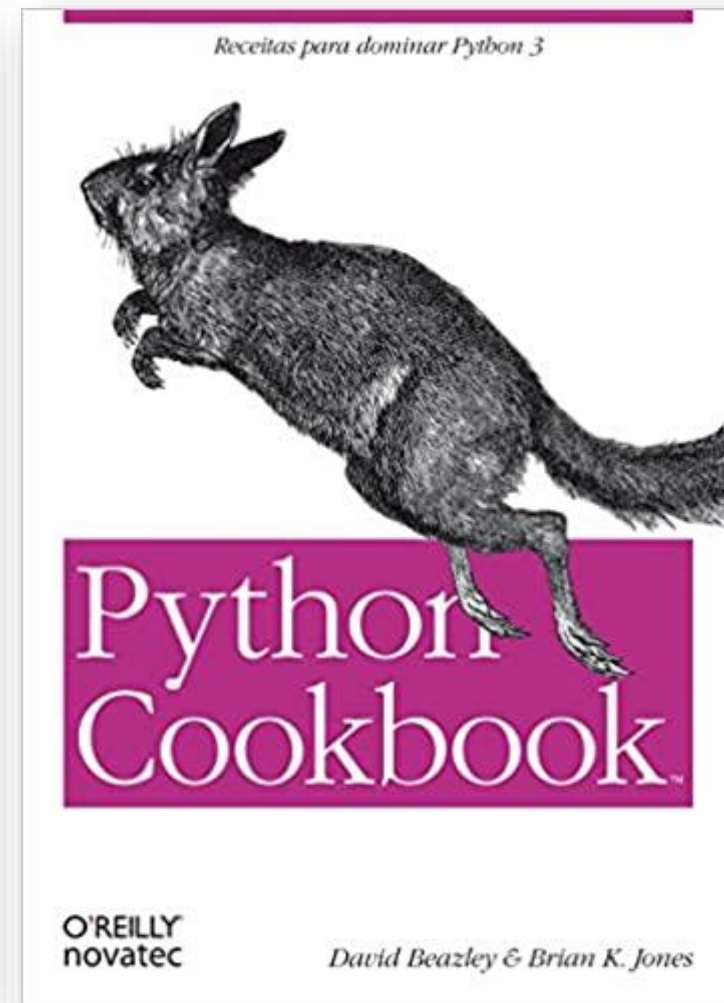
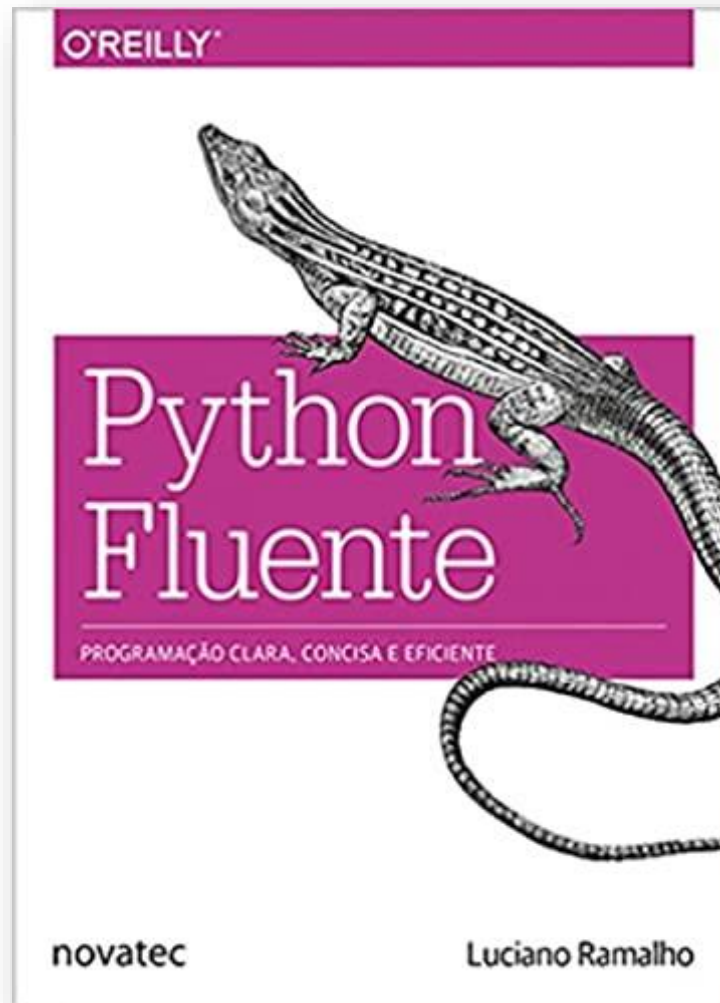
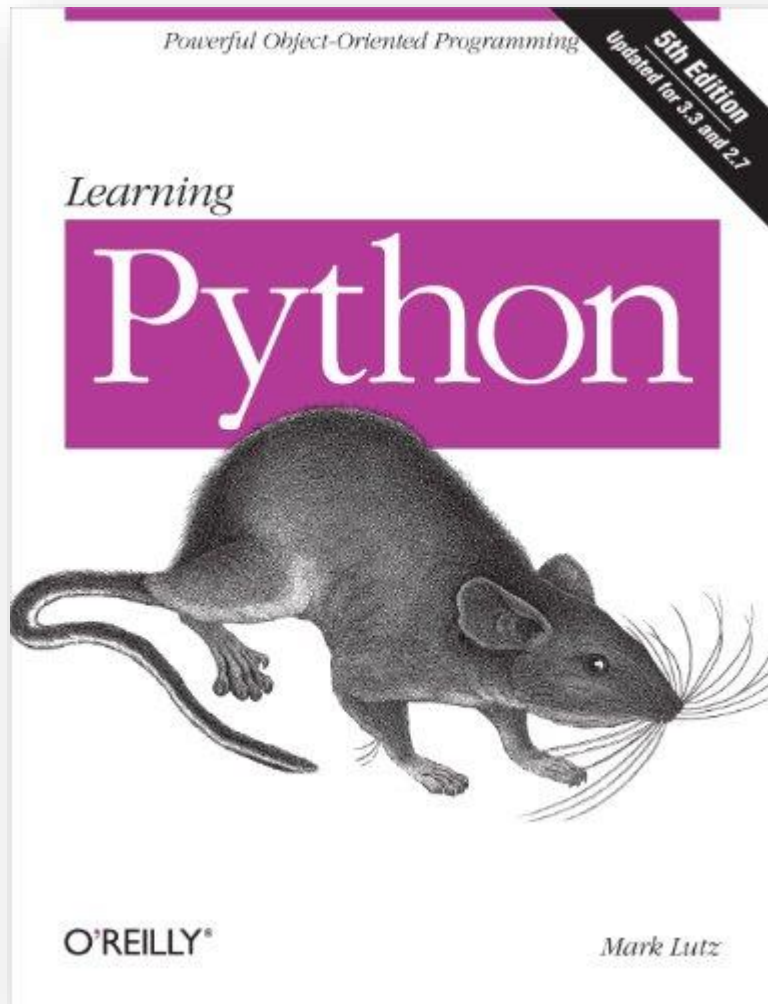
Referências Bibliográficas

LIVROS



Referências Bibliográficas

LIVROS



Referências Bibliográficas

LINKS, ÍCONES, IMAGENS

- As referências de links utilizados podem ser visualizados em <http://urls.dinomagri.com/refs>
- Tutoriais disponíveis no site oficial do Pandas - <http://pandas.pydata.org/pandas-docs/stable/>
- Livro de receitas disponíveis no site oficial do Pandas - <http://pandas.pydata.org/pandas-docs/stable/cookbook.html>
- As imagens foram Icon made by [Srip](#), [Pixel perfect](#), [Eucalyp](#) e [Prettycons](#) from www.flaticon.com

