

Conhecendo o AWS SageMaker para tarefas de Machine Learning

Cassiano Peres
Analista de Sistemas

Mais sobre mim

- CTO – Arabyka e Brexbit
- Graduado em TADS 2015
- Pós-graduando em Defesa Cibernética
- Apaixonado pela liberdade e descentralização
- Github: cassianobrexbit
- Linkedin: peres-cassiano

Desafio da Live

Hoje vamos explorar o AWS SageMaker, um serviço totalmente gerenciado da AWS, para preparar, criar, treinar e implantar modelos de Machine Learning (ML) de alta qualidade rapidamente.

Percurso

Etapa 1

Contexto de ML

Etapa 2

Explorando o AWS SageMaker

Etapa 3

Prática

Requisitos

- ✓ Conta ativa na AWS
- ✓ Fundamentos de ML
- ✓ Fundamentos de programação em Python
- ✓ Vontade e curiosidade

Machine learning

- IA e ML não são apenas sobre robôs e ficção científica
- Estão em todos os lugares, por exemplo:
 - Motor de buscas do Google
 - Sistema de recomendações da Amazon
 - Reconhecimento facial do Facebook.



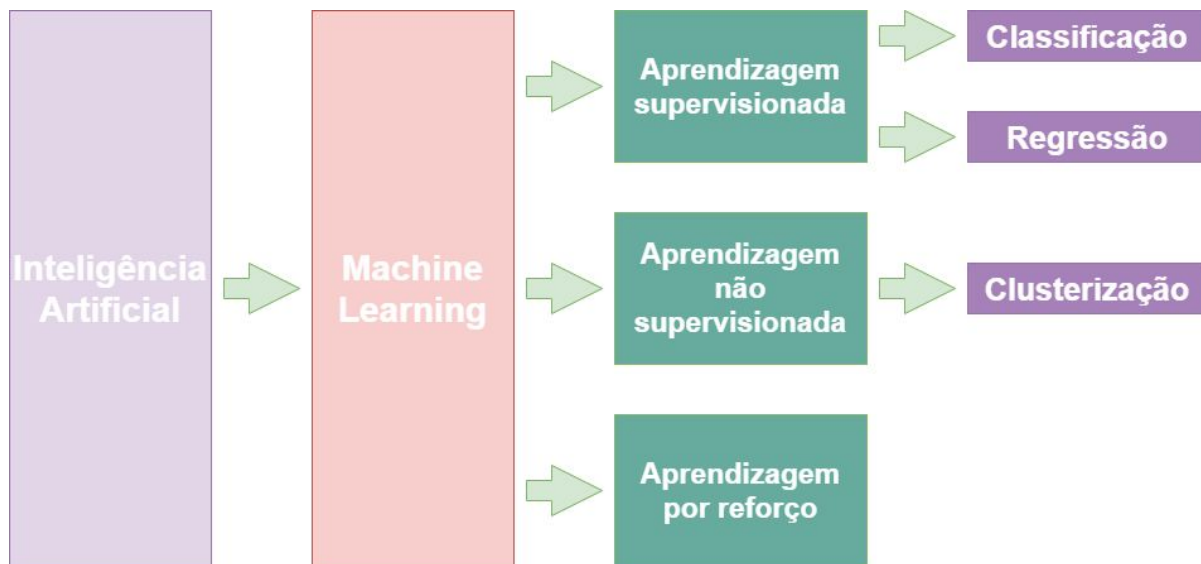
Machine learning

- Possuem três componentes básicos



Machine learning

- Panorama geral



Machine learning

- Good data vs bad data
- Good data
 - Grande volume de amostras
 - Não tendenciado
 - Sem dados incorretos ou faltantes
 - Contém dados relevantes apenas e não duplicados
- Bad data
 - Pouco volume de amostras
 - Tendenciado
 - Muitos dados incorretos ou faltantes
 - Muitos dados irrelevantes apenas e/ou não duplicados

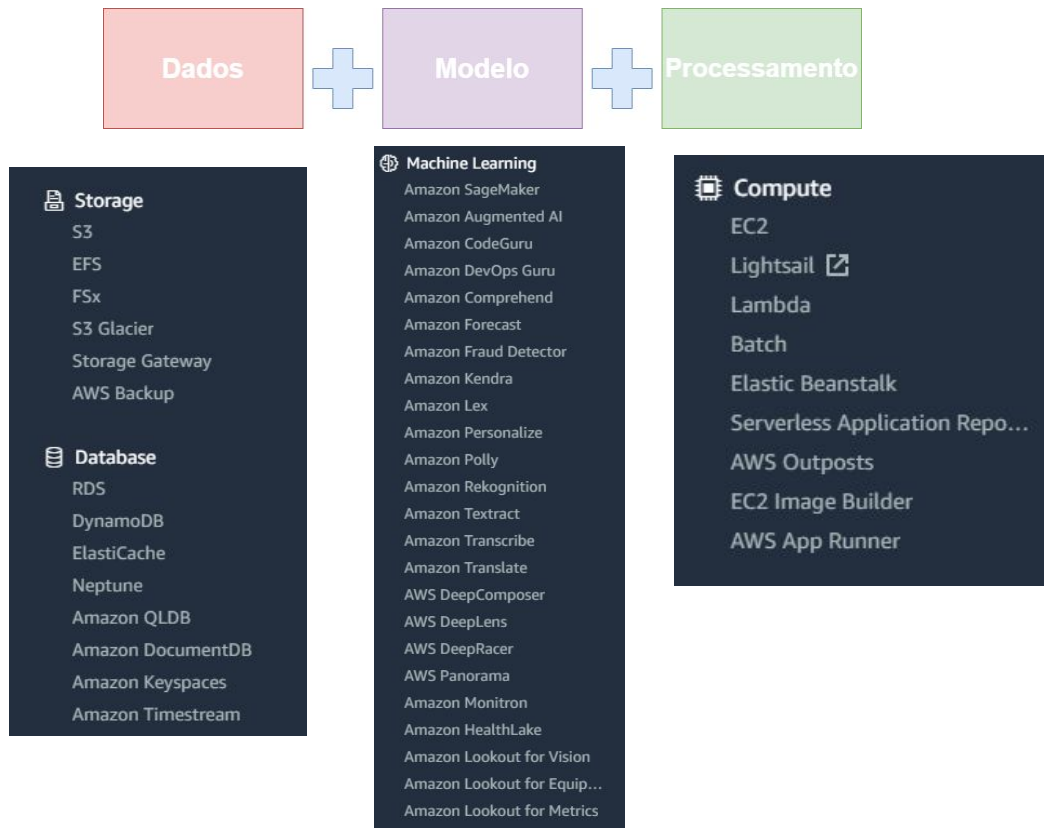
Machine learning

- Onde podemos encontrar dados?
 - Kaggle
 - UCI
 - AWS Dataset
 - ImageNet



DIGITAL
INNOVATION
ONE

Recursos de ML na AWS





AWS SageMaker

- Uma plataforma de fluxo de trabalho de ML totalmente gerenciada com serviços de rotulagem de dados, construção de modelo, treinamento, ajuste e implantação.
- Permite criar modelos escalonáveis de AI / ML com facilidade e eficiência.
- Permite criar modelos escaláveis de AI / ML com facilidade e eficiência.



DIGITAL
INNOVATION
ONE

AWS SageMaker





DIGITAL
INNOVATION
ONE

AWS SageMaker

Amazon SageMaker

Prepare →

SageMaker Ground Truth
Label training data for machine learning

SageMaker Data Wrangler NEW
Aggregate and prepare data for machine learning

SageMaker Processing
Built-in Python, BYO R/Spark

SageMaker Feature Store NEW
Store, update, retrieve, and share features

SageMaker Clarify NEW
Detect bias and understand model predictions

Build →

SageMaker Studio Notebooks
Jupyter notebooks with elastic compute and sharing

Built-in and Bring-your-own Algorithms
Dozens of optimized algorithms or bring your own

Local Mode
Test and prototype on your local machine

SageMaker Autopilot
Automatically create machine learning models with full visibility

SageMaker JumpStart NEW
Pre-built solutions for common use cases

Train & tune →

One-click Training
Distributed infrastructure management

SageMaker Experiments
Capture, organize, and compare every step

Automatic Model Tuning
Hyperparameter optimization

Distributed Training Libraries NEW
Training for large datasets and models

SageMaker Debugger NEW
Debug and profile training runs

Managed Spot Training
Reduce training cost by 90%

Deploy & manage →

One-click Deployment
Fully managed, ultra low latency, high throughput

Kubernetes & Kubeflow Integration
Simplify Kubernetes-based machine learning

Multi-Model Endpoints
Reduce cost by hosting multiple models per instance

SageMaker Model Monitor
Maintain accuracy of deployed models

SageMaker Edge Manager NEW
Manage and monitor models on edge devices

SageMaker Pipelines NEW
Workflow orchestration and automation

SageMaker Studio

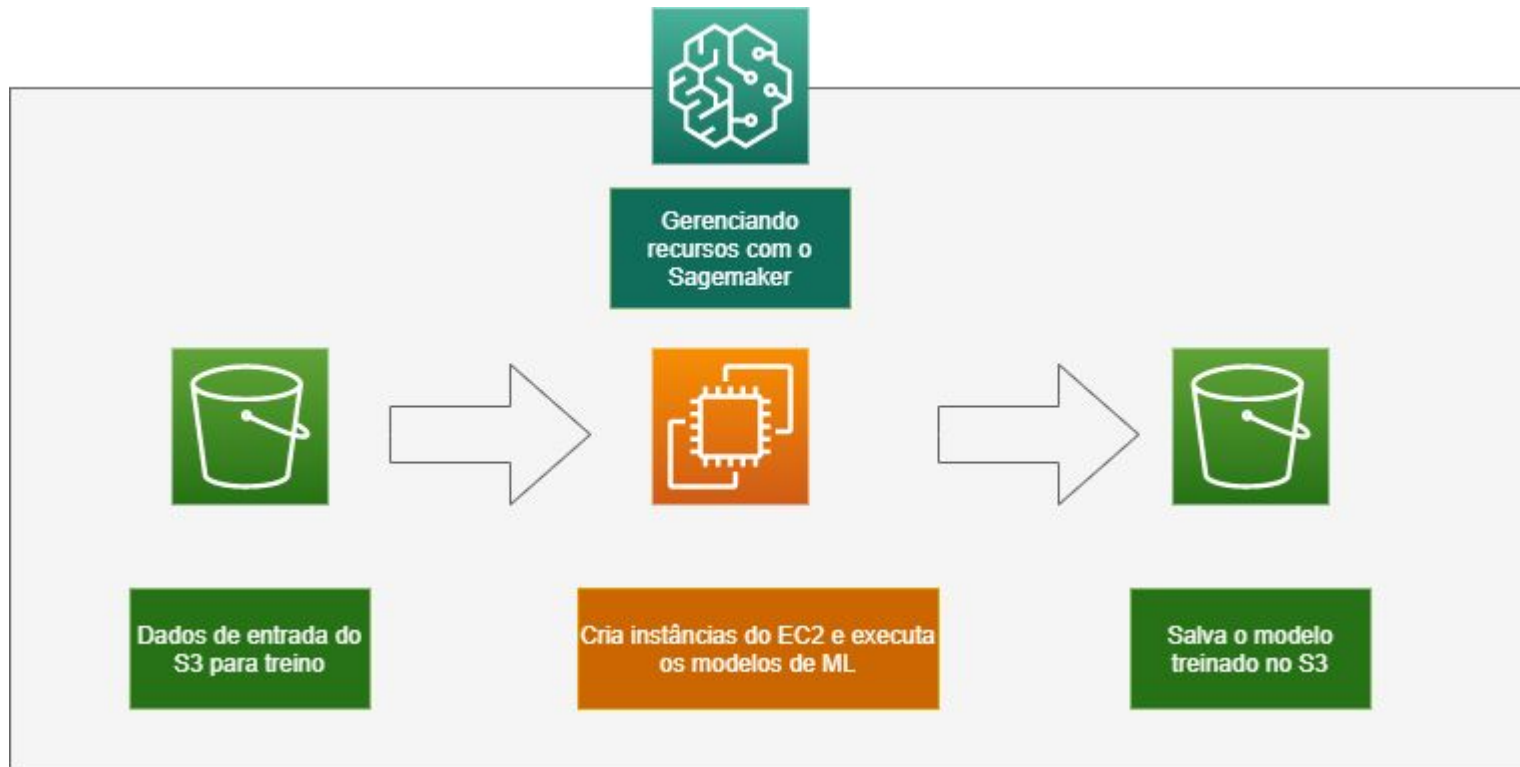
Integrated development environment (IDE) for ML

AWS SageMaker

- Dois componentes: Treinamento de modelo e deploy de modelo
- Para treinar um modelo de ML é necessário criar um job com os seguintes componentes
 - Bucket S3 com dados de treinamento
 - Recursos de processamento – Instâncias EC2
 - Bucket S3 para dados de saída
 - Caminho para o ECR onde estar armazenado o código de treino



AWS SageMaker



Desafio da Live

Nosso desafio é desenvolver um modelo de ML para prever se um cliente de um banco se inscreverá para um certificado de depósito (CD). O modelo será treinado no conjunto de dados de marketing que contém informações sobre a demografia do cliente, respostas a eventos de marketing e fatores externos.

Etapas

- Criar uma instância de bloco de anotações com Jupyter
- Preparar os dados
- Treinar o modelo para aprender com os dados
- Implantar o modelo
- Avaliar a performance do modelo ML

Obs: Os recursos utilizados estão dentro do nível gratuito da AWS.

Dicas de desafios

- Testar com outros modelos de ML
- Testar com outros tipos de problemas
- Utilizar outros conjuntos de dados
- Explorar os recursos do AWS SageMaker

Dúvidas?



Referencial

- <https://aws.amazon.com/sagemaker/>
- <https://aws.amazon.com/getting-started/hands-on/build-train-deploy-machine-learning-model-sagemaker/>
- <https://archive.ics.uci.edu/ml/datasets/bank+marketing>
- Repositório:
<https://github.com/cassianobrexbit/dio-live-sagemaker-16082021.git>