



Introdução ao Aprendizado de Máquina (Machine Learning) para Jornalistas de Dados



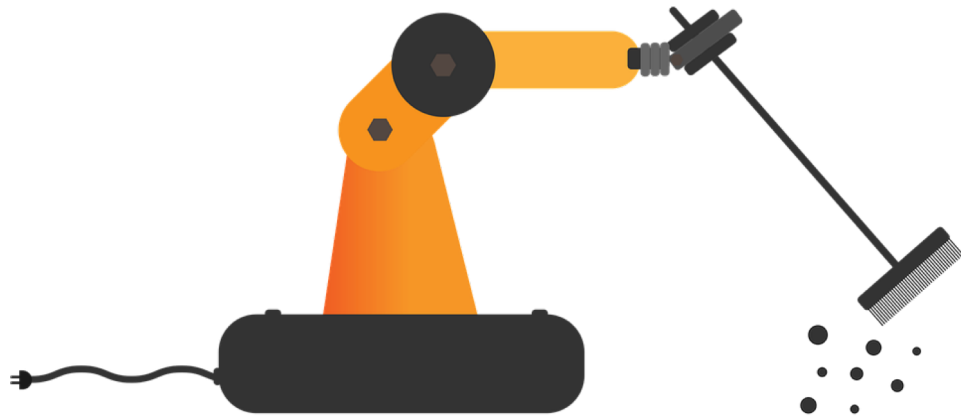
Dale Markowitz {Engenheira de IA Aplicada}

[@dalequark](#)

Google Cloud

O que é Aprendizado de Máquina (Machine Learning)?

automação?

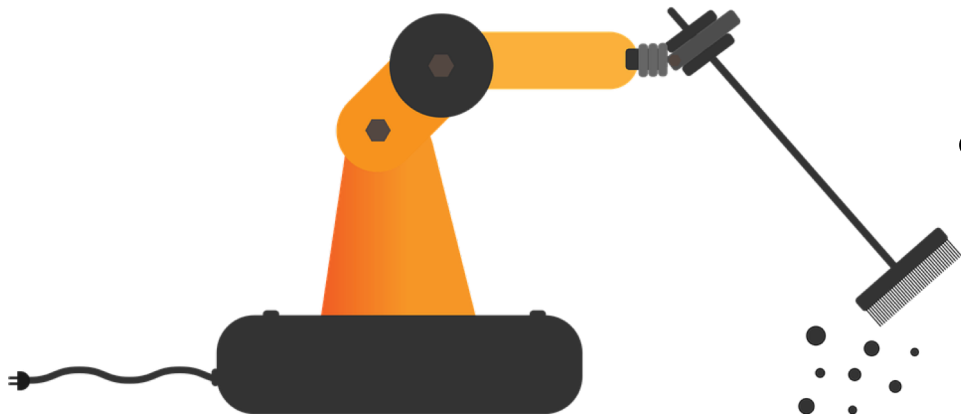




transcrição

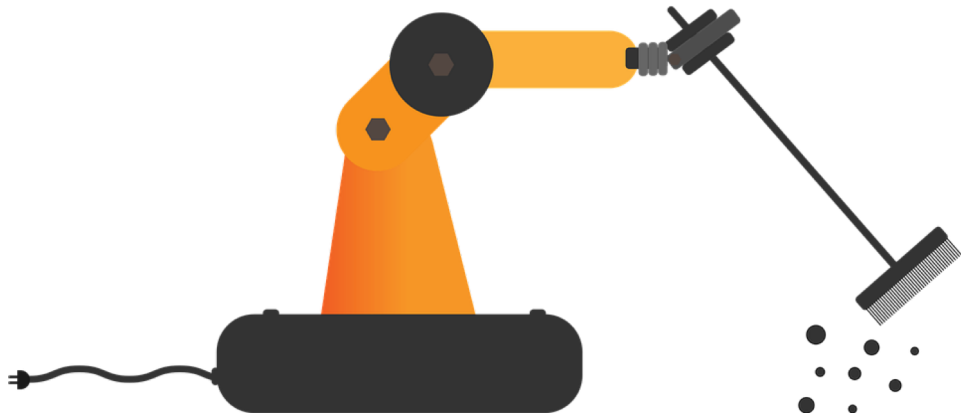


tradução

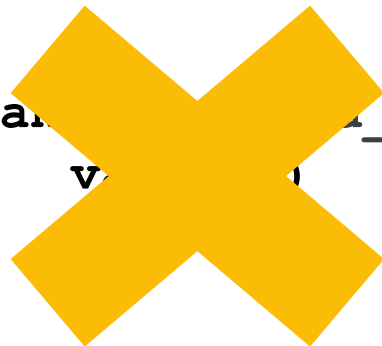


```
enquanto (energia_ligada) :  
    varrer ()
```

não é aprendizado de máquina



enquanto (você não está ligada) :



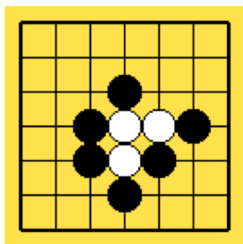
Recomendação



Previsão

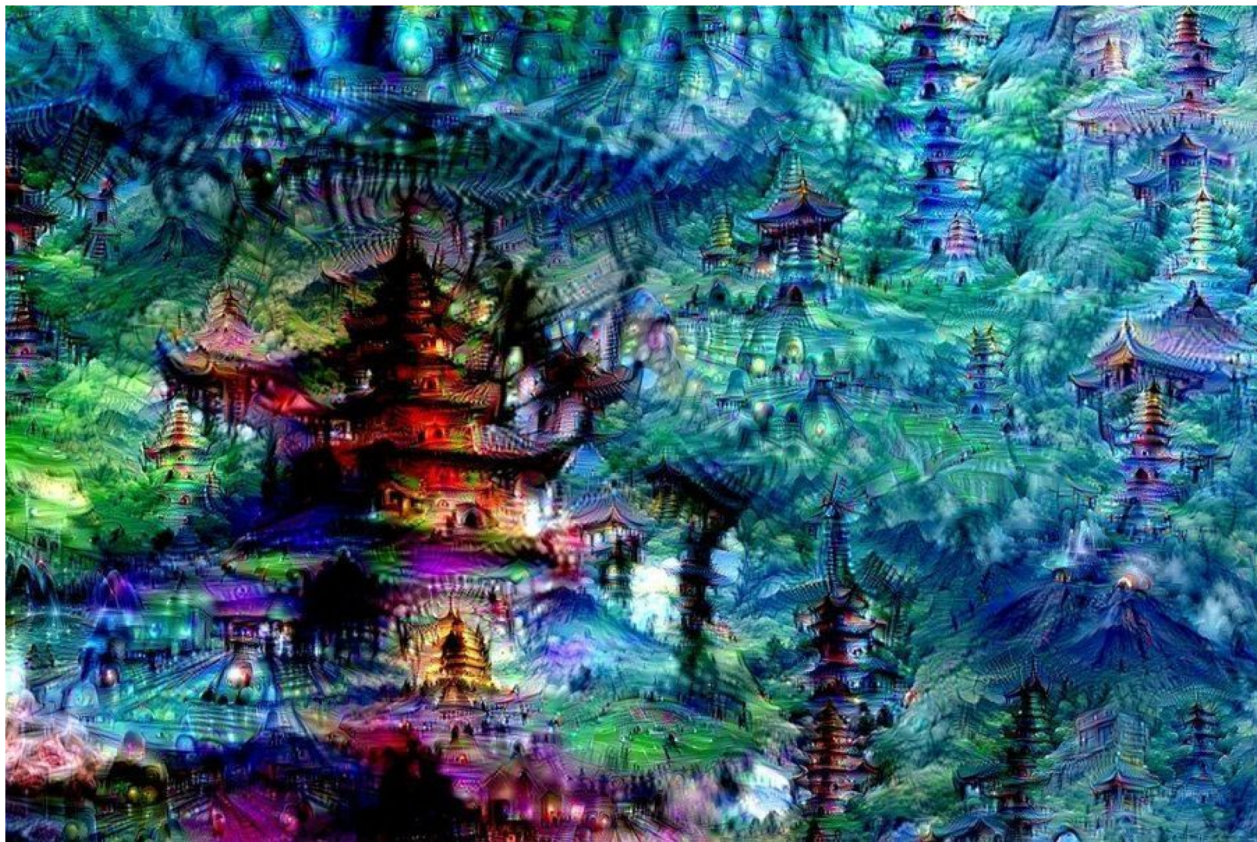


Estratégia de jogo

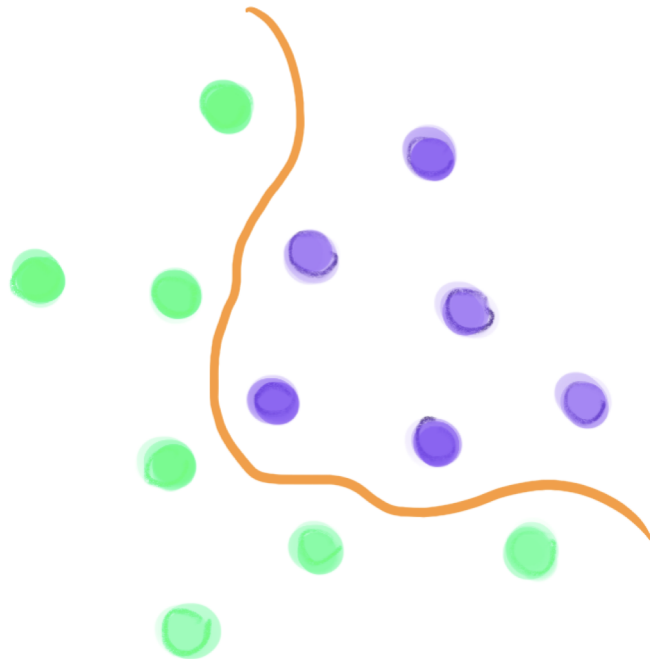


Classificação





Encontrando padrões em dados

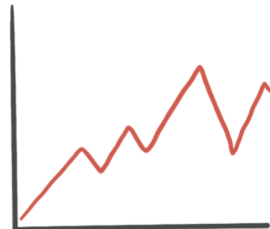


Um novo paradigma de programação

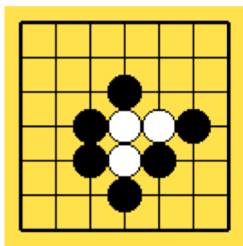
Recomendação



Previsão



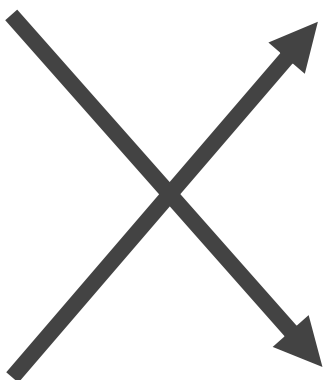
Estratégia de jogo



Classificação



classificação -> rotular coisas



Classificação

Cachorro

Gato

Programação Tradicional:

Programação Tradicional:

```
se (animal) :  
    tem uma cauda  
    longa  
    e  
    gosta de peixe  
    e  
    detesta  
    pessoas  
então:  
    retornar "gato"
```



Google Cloud

Programação Tradicional:

se (animal) :

tem uma cauda

longa

e

gosta de peixe

e

detesta

peessoas

então :

 **retornar** "gato"

regras

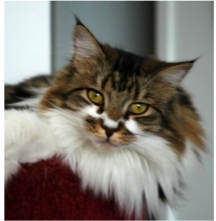
Aprendizado de Máquina:

Aprende por meio de **exemplos**, não **regras**

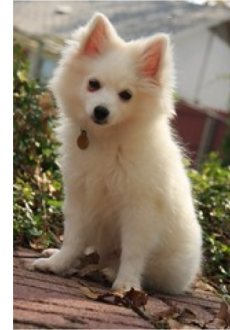
Construa um modelo

Conjunto de Dados para Aprendizagem de Rótulos

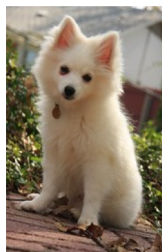
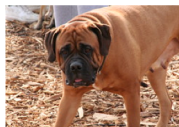
Exemplos de gatos



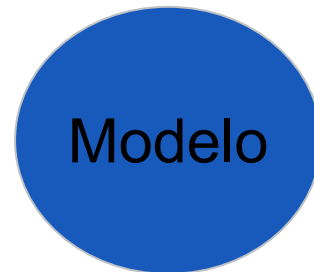
Exemplos de cachorros



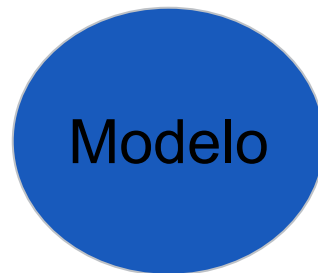
Treinando um Modelo



Algoritmo de
Aprendizagem
Automática



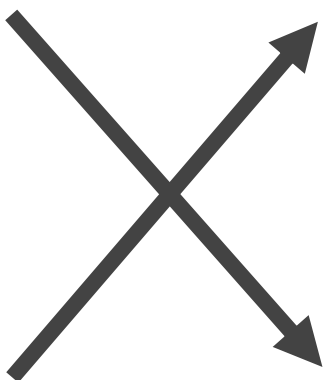
Fazendo previsões



“Cachorro”

Tipos de Aprendizado de Máquina

Aprendizado de Máquina para Fotos



Classificação

Cachorro

Gato

Saudável ou Pneumonia?



Tarefas comuns de Visão com Aprendizado de Máquina



rótulos

Landmark	96%
Place Of Worship	83%
Architecture	79%
Temple	74%
Hindu Temple	74%
Temple	72%
Tourist Attraction	71%
Building	68%
Amusement Park	64%



emoções

Joy		Very Unlikely
Sorrow		Very Unlikely
Anger		Very Unlikely
Surprise		Very Unlikely
Exposed		Very Unlikely
Blurred		Very Unlikely
Headwear		Very Likely

Roll: -2° Tilt: 4° Pan: 1°

Confidence 53%

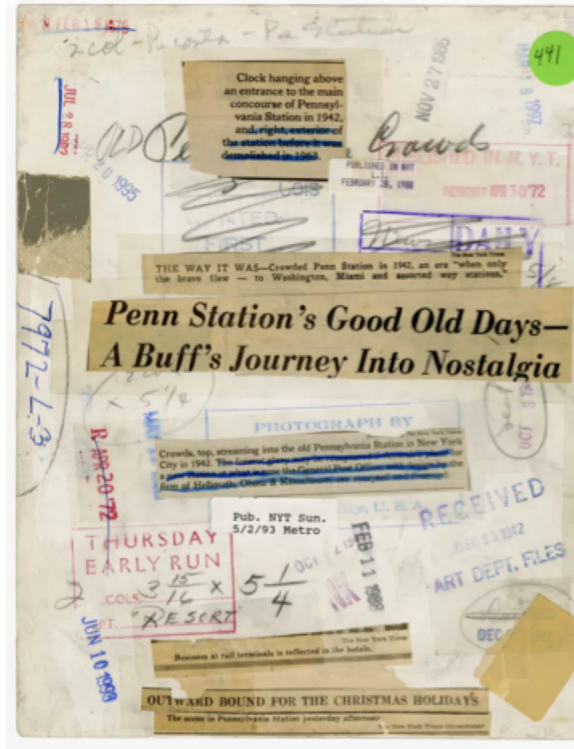


ocr
(reconhecimento
ótico de
caracteres)

“LUNA PARK”

Visão com Aprendizado de Máquina na Redação

arquivo digital do new york times





Times Open

[WHO WE ARE](#)

[CODE](#)

[DATA](#)

[PRODUCT AND DESIGN](#)

[WORKPLACE CULTURE](#)

[WORK WITH US](#)

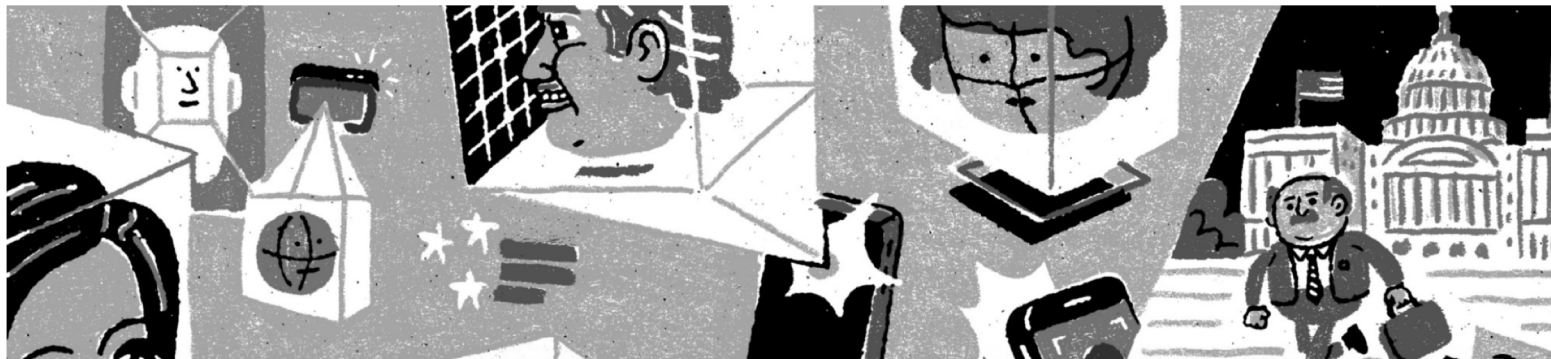
How The New York Times Uses Software To Recognize Members of Congress



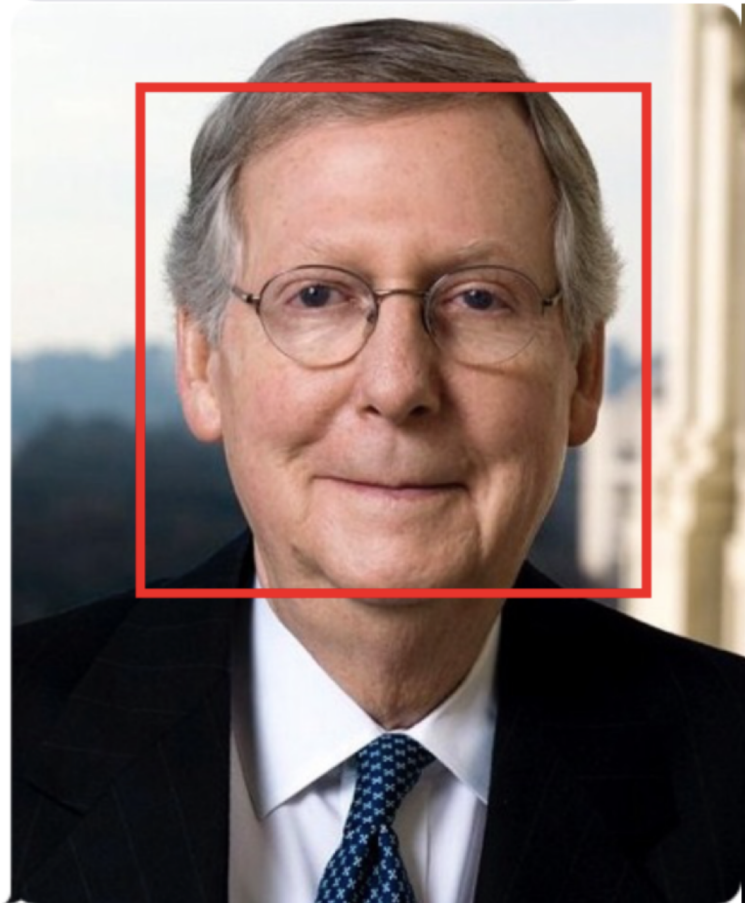
Jeremy Bowers

[Follow](#)

Jun 6, 2018 · 4 min read

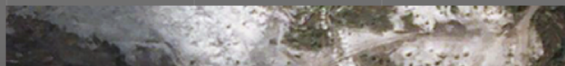


Mitch McConnell (red, almost certainly -
confidence = 100.0)

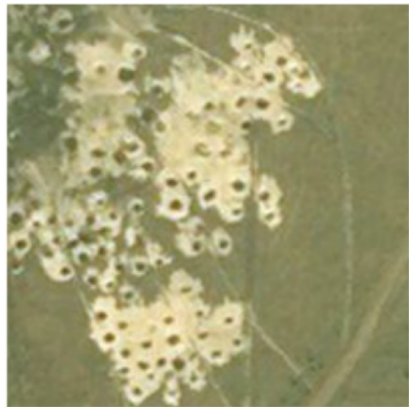
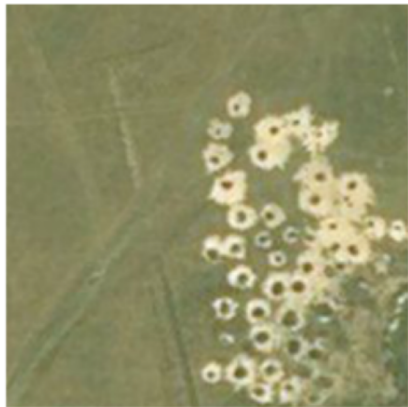
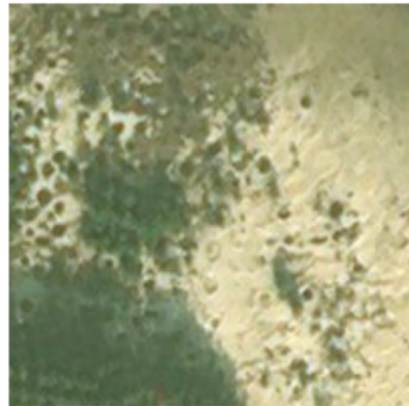
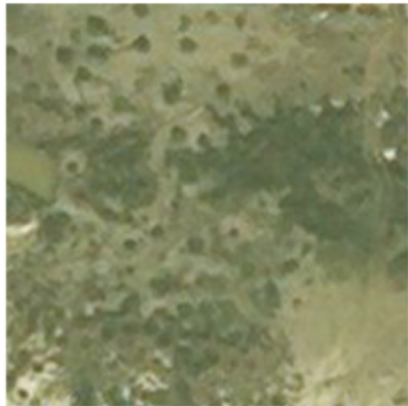


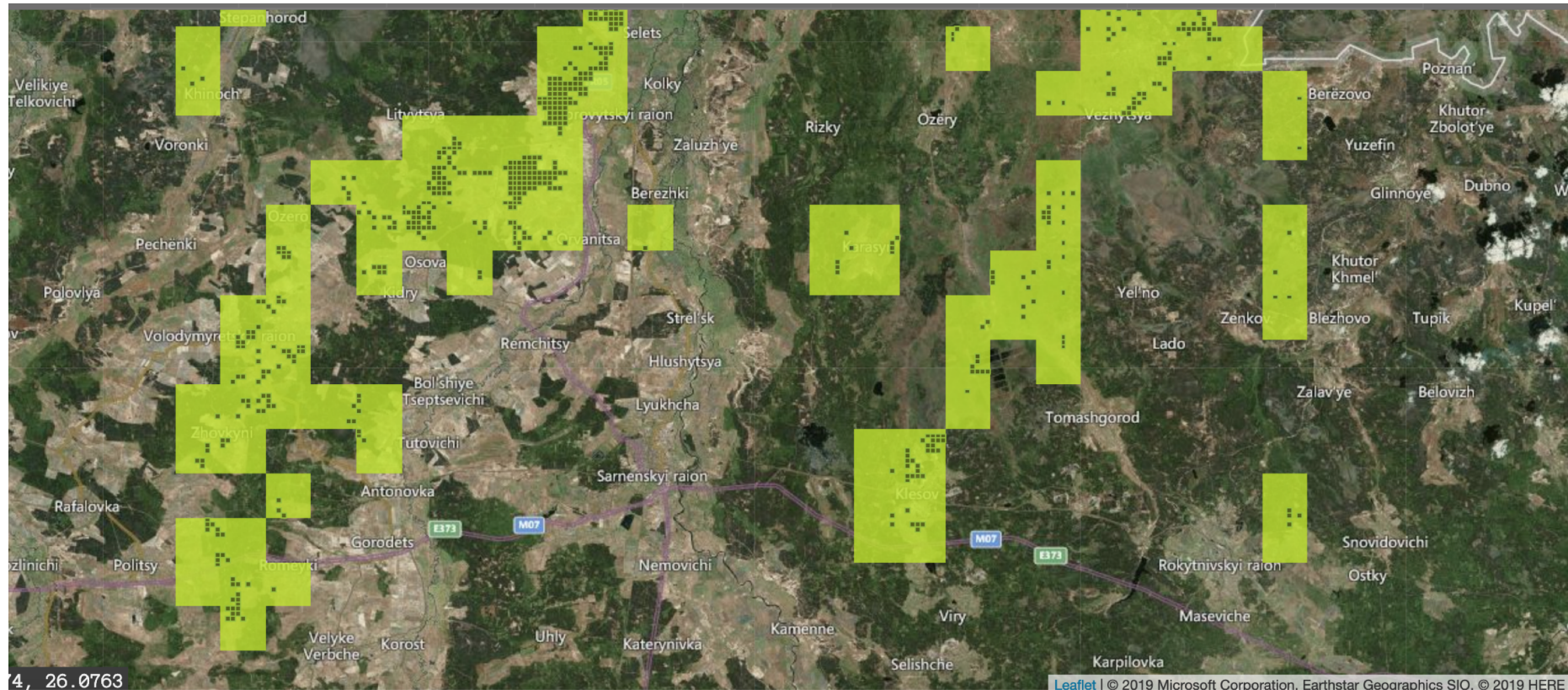
click me >

Leprosy of the land



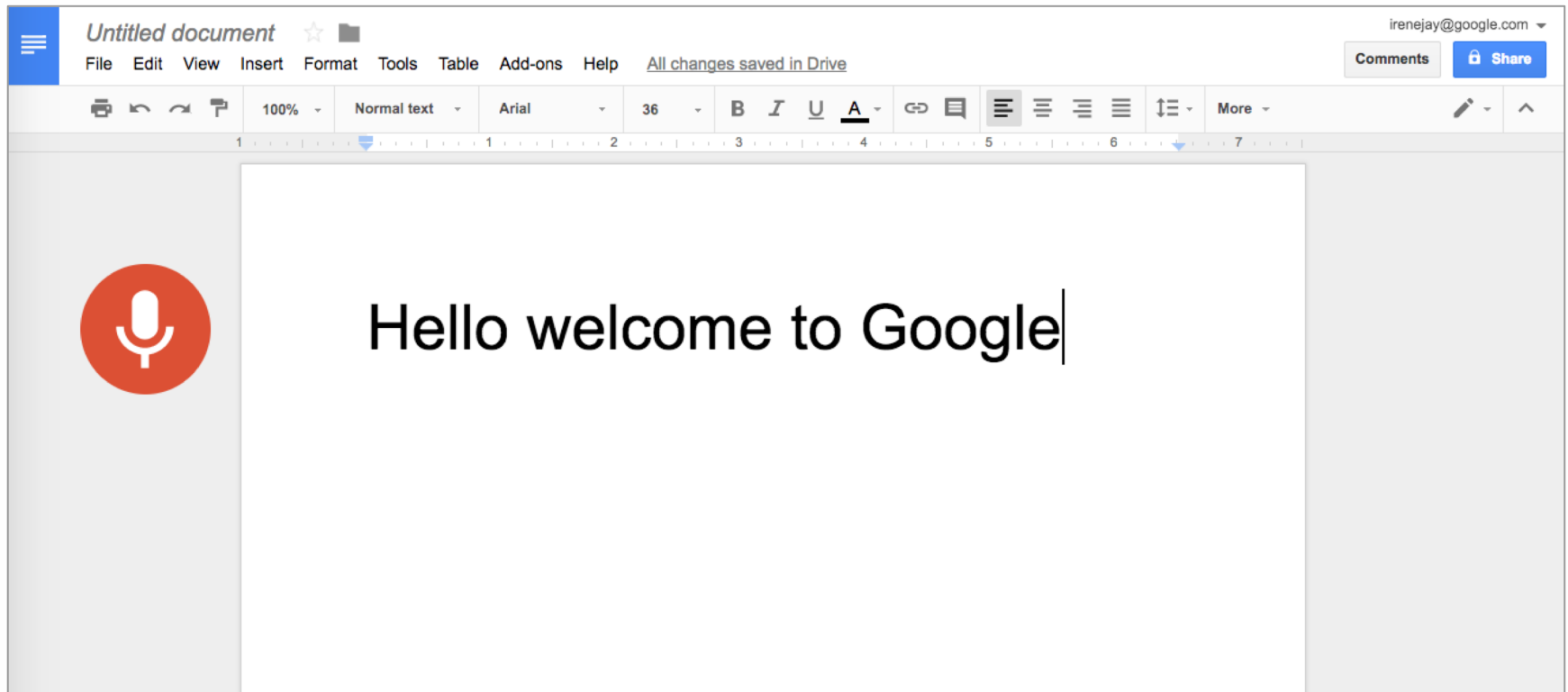
THOUSANDS HECTARES OF LAND TURNED





Aprendizado de Máquina para Áudio

Transcrevendo entrevistas

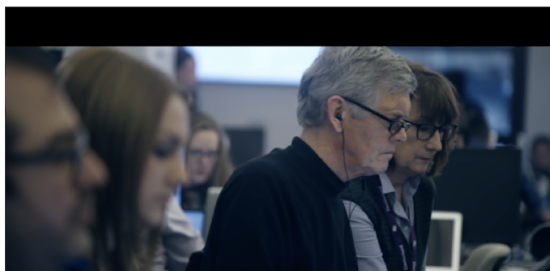


The image shows a screenshot of a Google Docs interface. At the top, the title bar reads "Untitled document" with a star icon and a folder icon. The menu bar includes "File", "Edit", "View", "Insert", "Format", "Tools", "Table", "Add-ons", and "Help". A status bar indicates "All changes saved in Drive". The user's email "irenejay@google.com" is visible in the top right corner, along with "Comments" and "Share" buttons. The toolbar contains various icons for document manipulation, including a microphone icon on the left. The main editing area displays the text "Hello welcome to Google" in a large, black, sans-serif font. The interface is clean and modern, with a light gray background and a blue header bar.

Google Cloud Voice API

Reduz o tempo de transcrição em 80%

The Washington Post



Name: Electionland_Cutdown_Master_061717.mp4

Format: mp4

Transcription Start Time: 2018-03-07 13:17:11

Transcription End Time: 2018-03-07 13:17:51

Transcription Duration: 0:00:40

most newsrooms ask a very important question on Election Day which is who won the great question we're asking a different question which is who can vote who can't vote where are there problems that are preventing people from being able to vote even though they're eligible the story of how voting actually happens at polling places in what the experience is like for the average motor that's a horror story to track so the idea behind election lamb was let's take advantage of technology and data to allow newsrooms and journalists to work together across the country to collaboratively cover the experience of voting on Election Day for 1100 journalists working either here New York at a local news rooms or in the journalism school so it's an enormous Coalition we're not covering politics for not covering the candidate this is my voting overall I think it's telling us that most of these incidents are isolated and journalist have been able to go and investigate cases and learned that goes problems are necessarily as widespread as we had beer there's been a lot of misinformation out there but we're dedicated to reporting out the facts and we're focused on making sure that the truth is what comes out in our reporting and then the recording of a partners

AI Watching Television News: Deep Learning Meets A Week Of Television

🕒 MAY 21, 2019

Television Explorer

What would it look like to have a computer watch a week of [television news](#) from CNN, MSNBC and Fox News, along with the morning and evening broadcasts of

Kalev Leetaru - @kalevleetaru

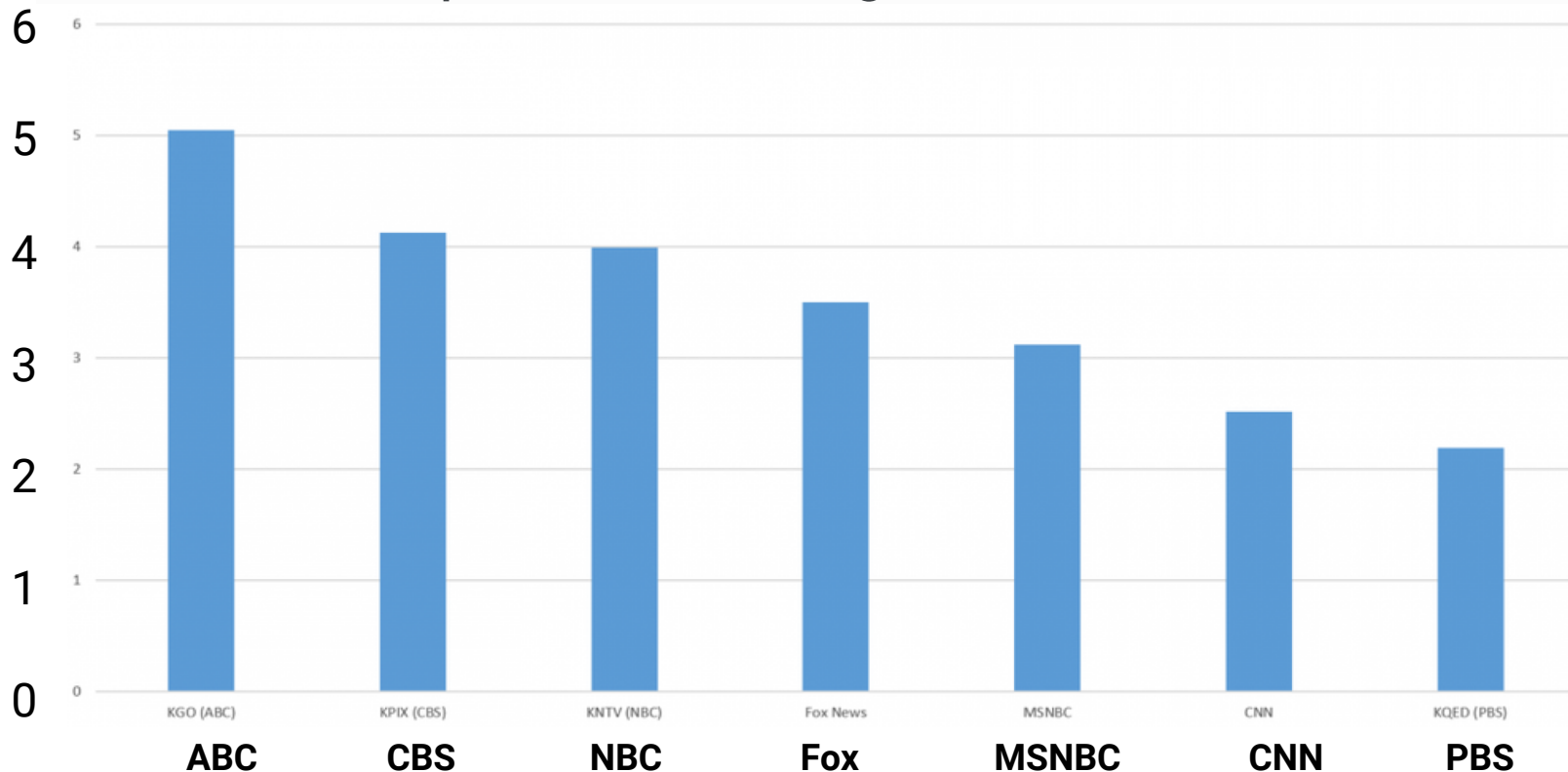
O que a IA (inteligência artificial) vê quando assiste a uma semana de notícias?

Usando Aprendizagem Profunda (Deep Learning) para Medir as Emoções Faciais da Televisão

A IA do Google Vision encontrou duas horas de tuítes de Trump em uma semana de notícias na TV

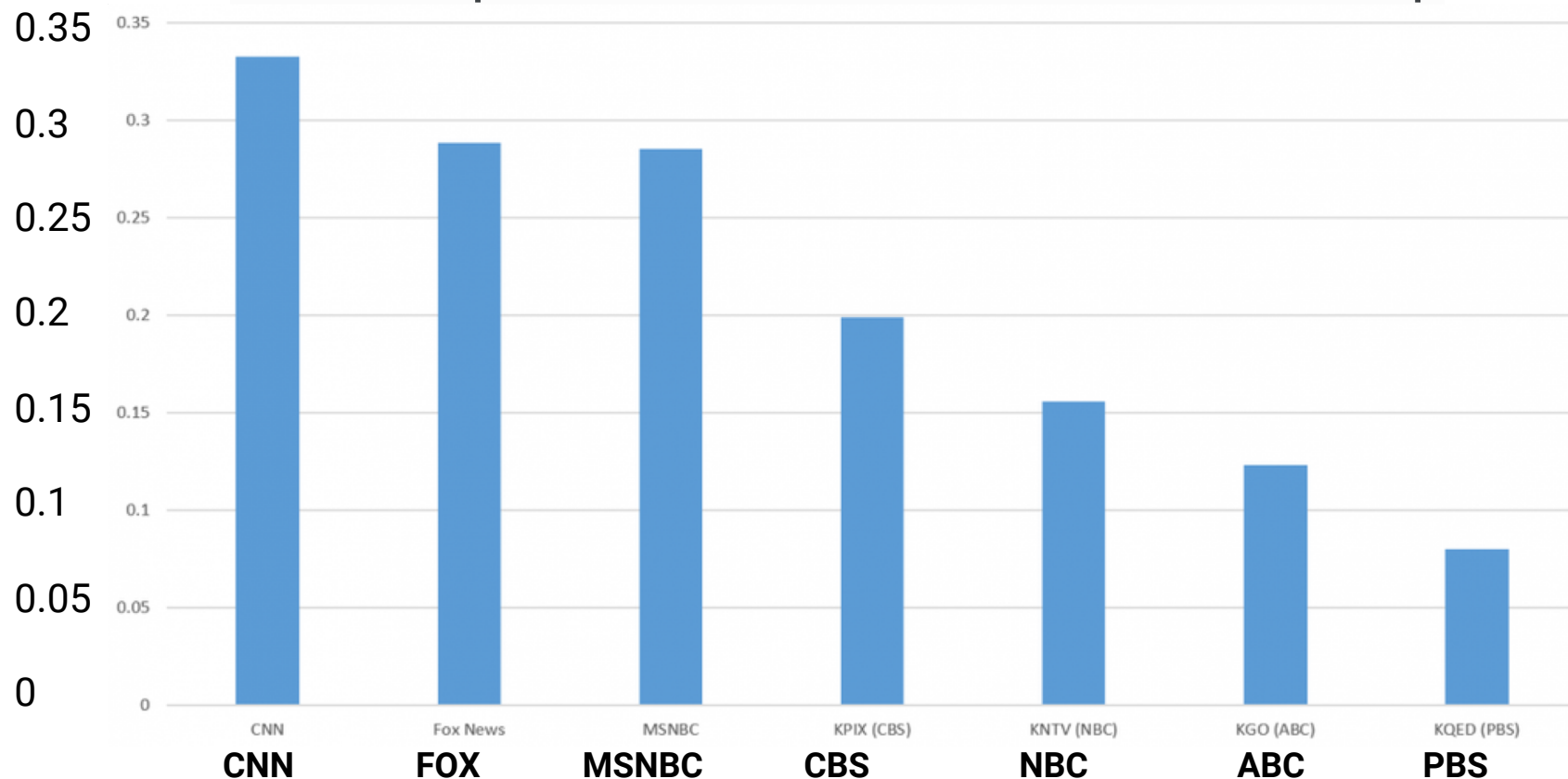
... e outros 13 artigos!

% de rostos expressando “alegria”, 15 a 22 de abril, 2019



Fonte: Forbes, Kalev Leetaru

% de tempo de TV mostrando tuítes do Trump



Fonte: Forbes, Kalev Leetaru

Aprendizado de Máquina para Dados de Texto

Classificação de Sentimentos



Frequent Flyer

@frequentflyer11

@Alta Fiquei parado esperando na pista durante horas!



Frequent Flyer

@frequentflyer11

@JetGreen tem tanto espaço para as pernas na classe econômica!



Classificação de Conteúdo



ADVERTISEMENT



CALIFORNIA

Times Investigation: LAPD misclassified nearly 1,200 violent crimes as minor offenses

By BEN POSTON, JOEL RUBIN AUG. 9, 2014 | 6:04 PM

Once police had Nathan Hunter in handcuffs, they tended to his wife.

She was covered in blood. She told the officers Hunter flew into a rage that night in February 2013 because she hadn't bought him a Valentine's Day gift. He beat and choked her before stabbing her in the face with a screwdriver and throwing her down a flight of stairs at their apartment in South L.A., according to police and court records.

ADVERTISEMENT



Aprendizado de Máquina para Dados Tabelados

(i.e. dados em planilhas)

BuzzFeed



BuzzFeed News Trained A Computer To Search For Hidden Spy Planes. This Is What We Found.

From planes tracking drug traffickers to those testing new spying technology, US airspace is buzzing with surveillance aircraft operated for law enforcement and the military.

Avião Espião ou Avião de Paraquedismo?

Quando usar Aprendizado de Máquina

1. Você consegue fazer sua tarefa manualmente?

2. Você tem dados categorizados?

3. Você precisa de explicabilidade?

4. Você pode tolerar erros?

Encontre as agulhas no palheiro

Ferramentas de Aprendizado de Máquina

DocumentCloud.org

Três Caminhos

Faça-Você-Mesmo

- Bibliotecas de Código Aberto em Python ou R (Gratuitas)
- Requer Experiência com Programação
- Requer Experiência com Ciência de Dados
- Requer Dados
- Funciona Localmente

Maior Customização

Cloud AutoML

- Requer Dados
- Requer Pouca ou Nenhuma Programação
- Serviço Pago
- Funciona na Nuvem

APIs de AM

- Usa Modelos Pré-Treinados (não há customização)
- Requer Alguma Programação
- Serviço Pago
- Funciona na Nuvem

Menor Customização

Use um **modelo pré-treinado** para cumprir tarefas comuns de AM



Cloud Vision



Cloud Video Intelligence



Cloud Speech-to-Text &
Text-to-Speech



Cloud Natural Language

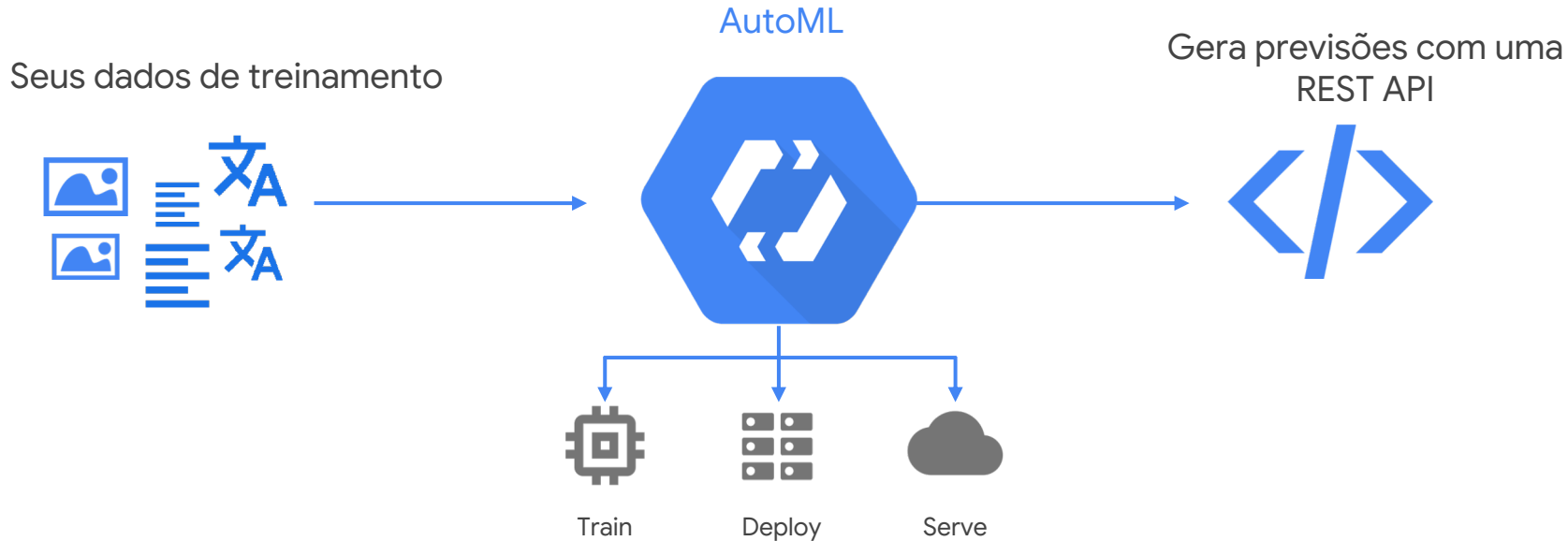


Cloud Translation



Cloud AutoML

O que é Cloud AutoML?





Vídeo



Vision



Tradução



Tabelas



Linguagem

Sumário

- O que é Aprendizado de Máquina?
- Como você pode usar Aprendizado de Máquina em reportagens?
- Como você *deveria* usar Aprendizado de Máquina?
- Ferramentas para começar