

# Utilizando Ipyannotator para anotações de tráfego terrestre

AUTHOR

Ítalo Epifânio

## Sumário

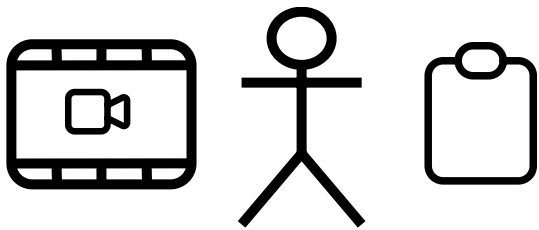
---

- Introdução a anotações
- Abordagem Data-centric AI e por que você deveria se importar
- Adicionando Ipyannotator ao seu workflow
- Analizando anotações de tráfego terrestre com Ipyannotator

## Processo de anotação

---

Entrada -> Manual, Simples & Processo Repetitivo -> Saída



Entrada -> Modelo ML -> Saída



ML Model



1. Data available can be analyzed and understood by humans but machines can't recognize such objects without training them with thousands of images that are annotated precisely
2. To analyze terrestrial data traffic a Machine Learning model previously trained with annotation data would receive an image as output and outputs objects that were recognized by the patterns find on the trained phase

## Fases de anotação

---

### 1. Anotação inicial

Tópico de pesquisa

Fonte: Tarefa relacionado, modelos pré-treinados, logs de usuários, manual

### 2. Supervisão

Humano ou validações semi-automáticas da rotulação

### 3. Monitoramento

Nem sempre executado. Importante para casos de uso reais.

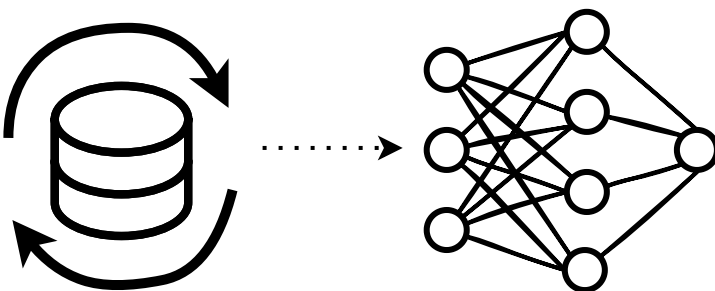
## Data-Centric AI

- A indústria está mudando de uma abordagem model-centric para uma data-centric
- Entender esse conceito pode ajudar a sua carreira e dar vantagem competitiva a sua empresa.

## Foco no dado

---

- É razoável
- Algumas vezes o dado é encarado apenas como uma entidade estática
- Melhorar a qualidade dos dados vai melhorar a performance da sua IA e reduzir BIAS
- Desenvolve métodos consistentes de coletar, rotular dados, assim como treinar, otimizar e atualizar modelos.



0. Vocês já viram algum cientista de dados dizer que um certo dado é bom?

1. AI is built with data and code, so data-centric it's reasonable, but not only practiced
2. Sometimes the data is something downloaded or used from another scientist research
3. BIAS is something very discussed in Brazilian articles about AI, understanding the data will help researchers to understand the data, check which resources need to be better described and improve the overall data quality
4. New methodologies are presented due to the quickly data iterations
5. Explain the image, discuss the data versioning and reproducibility challenges

## Erros de rotulação

Ferramentas como Cleanlab foram capazes de encontrar diversos erros em datasets considerados de boa qualidade

Milhões de errors foram encontrados <https://labelerrors.com/>

MNIST    CIFAR-10    CIFAR-100    Caltech-256    ImageNet    QuickDraw



given: 5  
corrected: 3



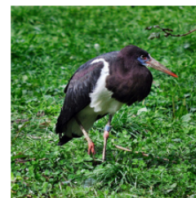
given: cat  
corrected: frog



given: lobster  
corrected: crab



given: ewer  
corrected: teapot



given: white stork  
corrected: black stork



given: tiger  
corrected: eye

E a qualidade dos seus dados?

1. Machine learning usually investigate, iterate and research about the model. The data is usually used as a static asset collected "before AI development". This approach lead IAs to a lot of bias and unexpected approaches.
2. Industry has been shifting for a more data-centric approach in the last few years. Models are more standardize and data hungry.
3. Training data, labeling, augmentation, slicing and management is the bottleneck and interface for today AI development.
4. Add example of current workflow[0] and mention that Ipyannotator can help developers to analyze the outputs.
5. Talk about manual labeling, the ethical and governance challenges

# Ipyannotator “The infinitely hackable annotation framework”

Um framework de anotação para o seu workflow de ML

## Motivação

---

- Lidar com o vasto número de tarefas de anotação, formato de dados e suas visualizações pode ser desafiador
  - Ferramental existente muitas vezes não é flexível o suficiente
  - Integração fácil com o ferramental do time de ciência de dados faz o trabalho ser mais rápido e reproduzível
  - Hackear as funcionalidades dos anotadores permite customizações e extensões
  - O ferramental como um todo está migrando para uma abordagem mais amigável a cientistas de dados (ex. PyScript)
1. Introduce Ipyannotator using the JOSS paper motivation 1.1. Talk about the literate programming style by mentioning the nbdev and other libs. This will attract the developer’s curiosity for the project.
  2. Talk about how Ipyannotator can help the data science and ML team to investigate the input, outputs, identify biases adding a workflow of constant iteration [0]

## Benefícios

---

- Ipyannotator fornece uma maneira de explorar e entender a qualidade dos seus dados
- A integração da biblioteca ao workflow do time de ciência de dados provê uma forma simples e barata de melhorar a qualidade de dados
- A criação de datasets com o ferramental que cientistas de dados já conhecem
- Funcionalidades facilmente customizáveis e extensíveis “infinitely hackable framework”

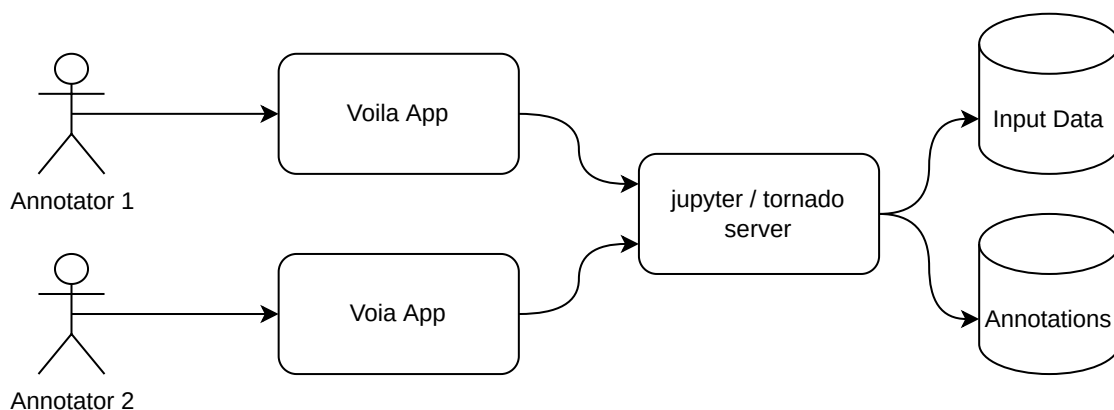
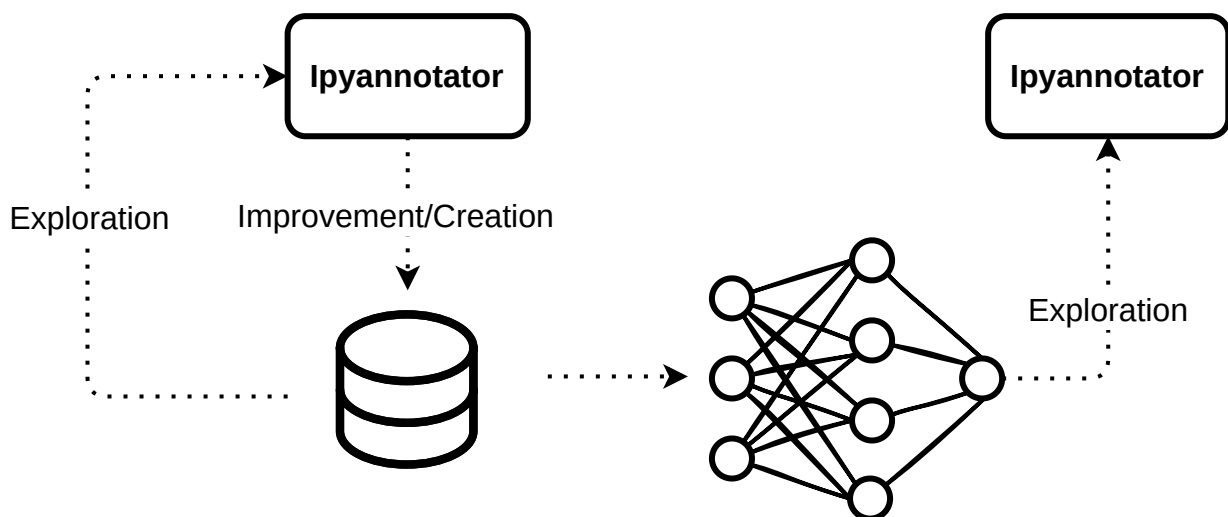
## Tecnologias

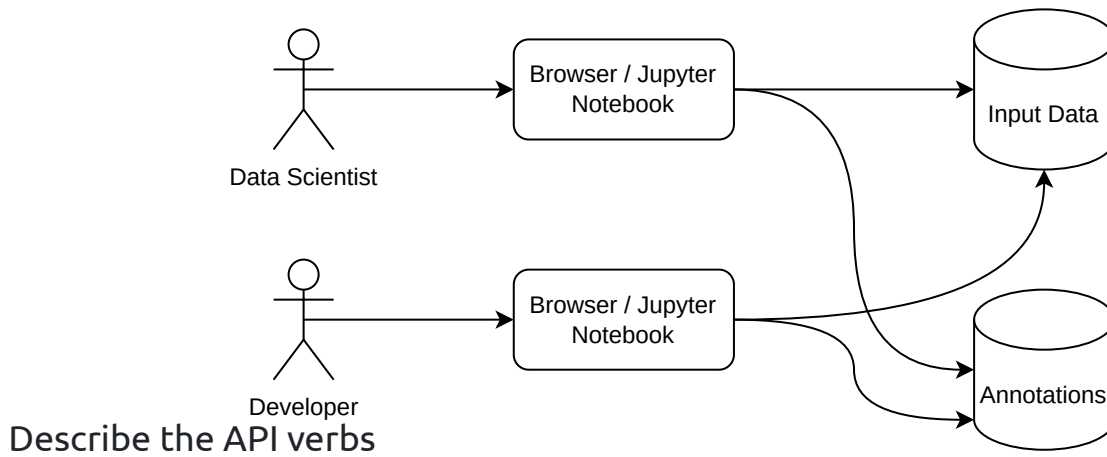
## Ipyannotator foi desenvolvido utilizando as seguintes tecnologias

- Nbdev
- Ipywidgets
- Ipycanvas
- Ipyevents
- Voila

Talk about each technology showing how it helps to build Ipyannotator ecosystem - Nbdev for build the lib using the Jupyter notebooks literate programming style - Ipywidgets for building the UI - Ipycanvas for user interactions with the UI - IPyevents for additional user interactions - Voila to execute a standalone web application

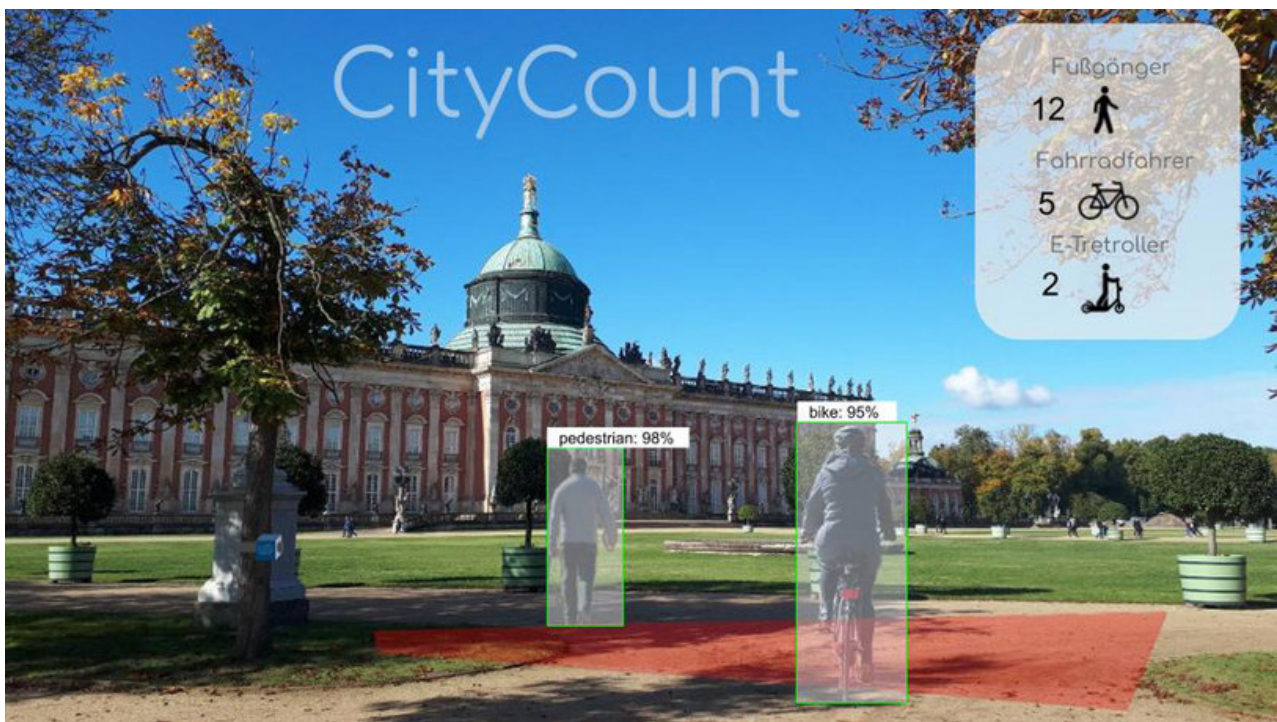
## Arquitetura e uso





# Anotações de tráfego terrestre

## CityCount Project



- Controle e gerenciamento de tráfego terrestre
- Soluções orientada por dados para mudanças de tráfego no centro da cidade
- Projetar uma demonstração inicial de um dispositivo móvel para planejamento de tráfego
- Nenhum dado pessoal é armazenado

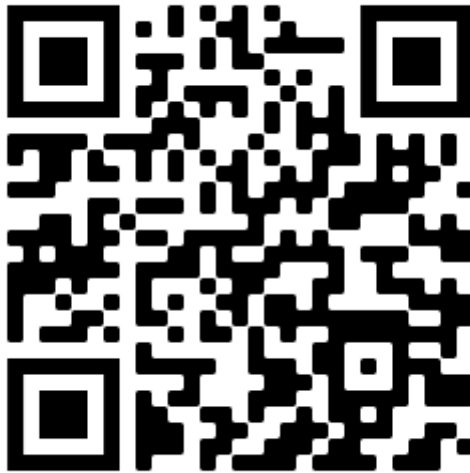
## Exemplos visuais

(Um momento)

## Open Source Software

---

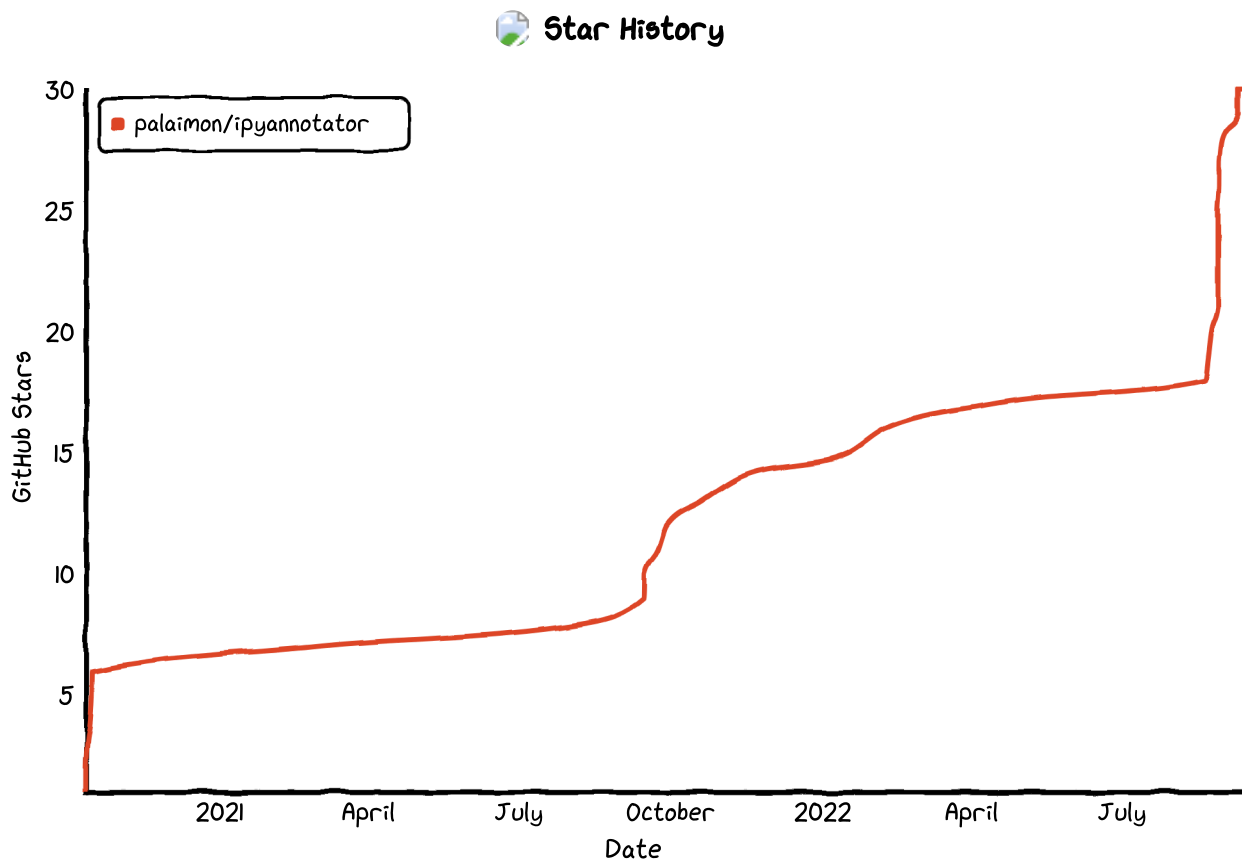
- <https://github.com/palaimon/ipyannotator>
  - Dê uma estrelinha
  - [Artigo JOSS](#)
  - Abram issues para perguntas, bugs e ideias



Live star

---





Star History Chart

## Obrigado pela atenção!

---