



**GAIA**  
powered by  **BNDES**

# Time 09

---



**Pedro Andrade**, cientista de dados na segov-mg, e montanhista e corredor nas horas vagas :)



**Renata Guanaes**, servidora pública federal mergulhada na Ciência de Dados, apaixonada por livros, corrida e tênis.



# Problema

**BBC NEWS | BRASIL**

Os 15 países que emitiram mais CO2 nos últimos 20 anos (e em que posição está o Brasil)

21 dezembro 2019



**Emissão de CO2 segue em nível recorde mesmo com a pandemia, diz ONU**

De acordo com o boletim anual, a concentração de dióxido de carbono na atmosfera aumentou dramaticamente em 2019 e continuou em 2020

**BBC NEWS | BRASIL**

**Desmatamento avança no Cerrado e faz crescer risco de apagão elétrico no Brasil**

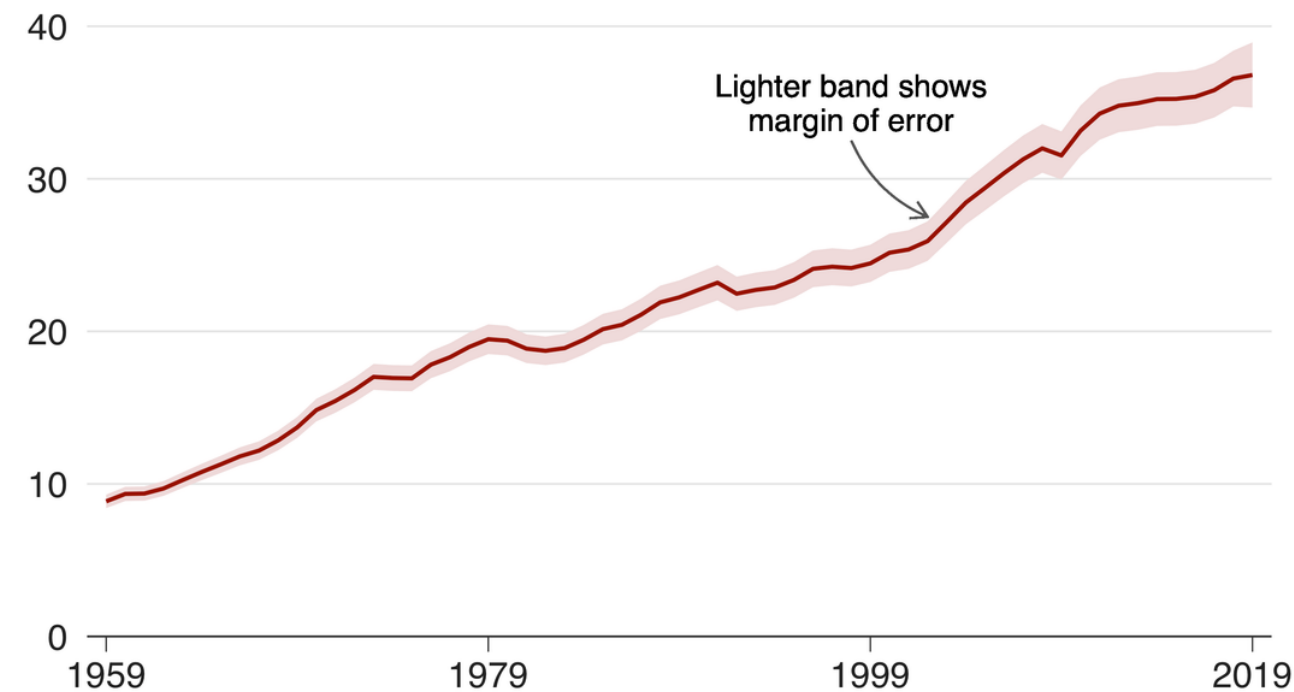
**CartaCapital**

SUSTENTABILIDADE

**Queimadas e desmatamento na Amazônia puxam aumento das emissões brasileiras**

## Global CO2 emissions continue to rise

Emissions in gigatonnes of carbon dioxide



Source: Global Carbon Project/CICERO

**BBC**

**GAIA**  
powered by  **BNDES**

# Problema

## Forbes

**Banco Central pressiona por bancos mais verdes**



Caso incorpore os riscos relacionados às mudanças climáticas nos testes de estresse, instituição brasileira pode garantir a vanguarda dos reguladores em todo o mundo

O Banco Central está se preparando para exigir que os bancos contabilizem potenciais perdas advindas de fenômenos relacionados a **mudanças climáticas**, como secas, inundações e incêndios florestais, posicionando-se como líder global na regulamentação do setor financeiro com base em ESG.

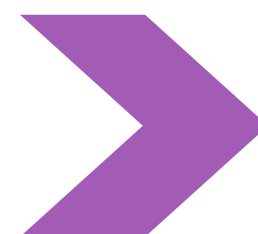
**GAIA**  
powered by  **BNDES**

# Oportunidade

<https://www1.folha.uol.com.br/mercado/2021/06/entenda-o-que-e-esg-e-por-que-a-sigla-virou-febre-no-mundo-dos-negocios.shtml>



**Ambiental**



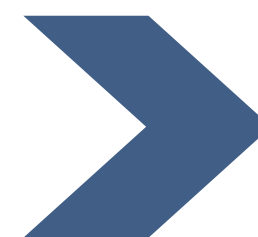
**é ser uma empresa que cuida do meio ambiente e preserva o planeta**

**Social**



**é ser uma empresa que implementa políticas de diversidade e promove impacto social positivo**

**Governança**

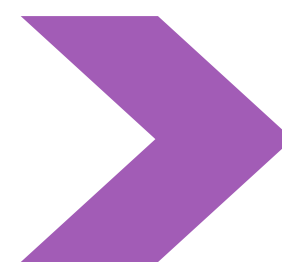
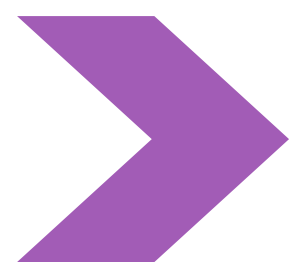


**é ser uma empresa que proporciona transparência fiscal e combate à corrupção**

# Oportunidade

**Desenvolvimento**  
Nosso foco principal

- Banco de Impacto
- Banco Social
- Banco Sustentável



**BNDES Verde:**

- preocupação com meio ambiente
- fomento aos investimentos sustentáveis



**GAIA**  
powered by  **BNDES**

# Problema no contexto do BNDES

---

- Como mensurar riscos ambientais das operações diretas e indiretas do BNDES?
- Como saber se a carteira de clientes da instituição está exposta a estes riscos ambientais?
- Como que o BNDES pode verificar a relação das suas operações de financiamento com:
  - a **emissão de carbono**,
  - a **produção de poluentes nocivos** para o meio ambiente,
  - a **geração de resíduos**,
  - o **consumo de água**,
  - a **preservação das unidades de conservação**,
  - a **manutenção das terras indígenas**
  - e a ocorrência de **desflorestamentos**?
  - e tantas outras informações?
- E como que todas essas verificações poderão ser feitas de forma **integrada, fácil, eficiente e transparente**?



# O que nos motiva

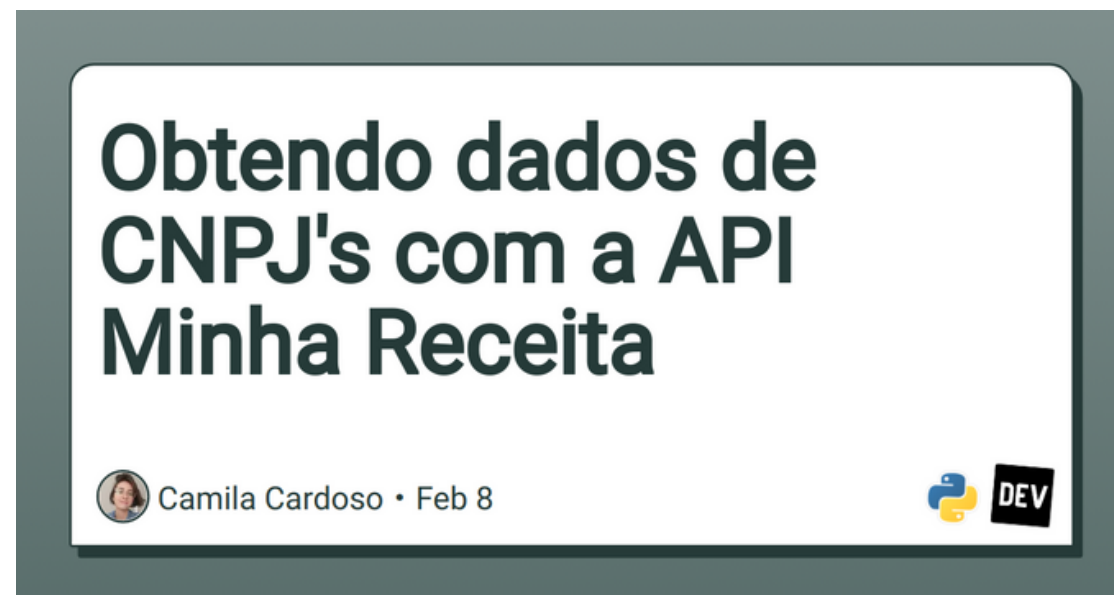


- Esse problema nos motivou a gastar nosso tempo, energia e paixão por inovação em uma solução que possa servir não só ao **BNDES**, mas a todos os **interessados** nos riscos ambientais da carteira de clientes do banco, em nível setorial e geolocalizado;
- Por isso, concebeu-se o **Gaia**, um modelo que para mensurar **os riscos de determinado CNPJ possuir atividades potencialmente poluidoras**;
- O modelo foi projetado de forma a utilizar apenas informações do próprio cartão de CNPJ para mensurar o risco daquela empresa ser poluidora e assim ser um projeto que se replique para outras áreas



# Base de dados - Modelagem

- Utilizou-se *datasets* com dados de CNPJ:
  - CNPJ na base de dados do CTF/APP (\*);
  - Demais CNPJs na base de dados da Receita Federal (*que não constam na base acima*).
- Para buscar os dados de informações de todos os CNPJs (*até daqueles presentes na base de operações indiretas do BNDES*) fez-se requisições na API do projeto *open-source* "Minha Receita" (\*\*);
- Informações coletadas dos CNPJs: código do município, porte da empresa, CNAE primária e secundárias, valor do capital social, entre outros.



(\*) Cadastro Técnico Federal de Atividades Potencialmente Poluidoras e/ou Utilizadoras de Recursos Ambientais.

(\*\*) Com a abordagem do artigo da Camila Cardoso.

# Base de dados - Análise

Além das bases de dados úteis para modelagem, utilizou-se, para análise, as demais bases listadas abaixo:



- Emissão de Poluentes Atmosféricos (IBAMA)
- Acidentes ambientais (IBAMA)
- Áreas embargadas (IBAMA)
- Unidades de conservação (IBAMA)
- Ocorrências de Incêndio Florestais (IBAMA)
- Reservas indígenas (FUNAI)
- Geolocalização Municípios Brasileiros ([github.com/kelvins](https://github.com/kelvins))
- PIB Municipal (IBGE)

# Ferramentas de Modelagem

- Para modelagem, optou-se por **ferramenta de automação de machine learning** desenvolvido pelo Sber Bank AI Lab, que reduz o tempo gasto neste processo e permite que a equipe se debruce sobre outros aspectos da solução;
- Utilizou-se **redes neurais** como algoritmo de cálculo da curva para mensuração do risco;
- Para **processamento de textos**, utilizou-se do **modelo pré-treinado "bert-base-multilingual-cased"**, desenvolvido pelo time de research do google e disponibilizado no HuggingFace.

sberbank-ai-lab/  
**LightAutoML**

LAMA - automatic model creation framework



10  
Contributors

4  
Issues

522  
Stars

53  
Forks



GAIA  
powered by BNDES

\* Nota: Mais informações de parâmetros do modelo consta no repositório em um notebook.



# Output e desempenho do modelo

- O modelo **Gaia** gera como saída **a probabilidade de determinado CNPJ** ser uma empresa poluidora (entre 0 e 1): quanto maior essa probabilidade, mais a empresa tende a ser poluidora;
- Além disso, o modelo considera como poluidora as empresas com **probabilidade maior que 0.5** (representado pela string "s"), e as demais como não potencialmente poluidoras (representado pela string "n");
- As probabilidades e o output do modelo constam, respectivamente, nas seguintes colunas do *dataset*: **"cnpj\_polui\_prob"** e **"cnpj\_polui"**;
- As **métricas** abaixo atestam a **boa performance deste modelo de classificação**, que possui além de acurácia alta, um bom balanceamento entre os dois resultados.

	precision	recall	f1-score	support
n	0.91	0.92	0.91	9247
s	0.87	0.85	0.86	5767
accuracy			0.89	15014
macro avg	0.89	0.89	0.89	15014
weighted avg	0.89	0.89	0.89	15014

\* Nota: Mais informações de desempenho do modelo consta no repositório em um notebook.

# GAIA.dash

---

- Para analisar os resultados do modelo, **criou-se um dashboard** com as seguintes funcionalidades:
  - Análise Exploratória dos Dados sobre de dados de **financiamentos diretos e indiretos**;
  - Análise Exploratória dos Dados sobre os **resultados gerados pelo modelo Gaia**;
  - Painel de consulta dos **riscos ambientais, por município e CNAE, sobre operações indiretas automáticas**;
  - **Mapas interativos** sobre:
    - Distribuição espacial das operações de financiamento, sejam poluidoras ou não, em comparação com a **emissão de carbono**;
    - Distribuição espacial das operações de financiamento, sejam poluidoras ou não, com **áreas de Unidade de Conservação, Terras Indígenas e Áreas Embargadas**;
    - Distribuição espacial das operações de financiamento, sejam poluidoras ou não, com **pontos de acidentes ambientais e incêndios florestais**.

# GAIA.dash

Visite o dash, interaja e dê feedbacks!



[https://hipotumos.shinyapps.io/gaia\\_bndes/](https://hipotumos.shinyapps.io/gaia_bndes/)



GAIA.dash

Sobre o projeto

AED - Financiamentos

AED - Resultados GAIA

Consulta - Resultados GAIA

Por CNAE

Por Município

Mapas

Consulta - Potencial de empresas poluidoras por CNAE

Selecione CNAE

Comércio varejista de produtos alimentícios em geral ou especializado em produtos alimentícios não especificados anteriormente

270

Operacoes com empresas potencialmente poluidoras

490

Operacoes com empresas nao potencialmente poluidoras

0.427952130968421

MÃ@dia das probabilidades do CNAE selecionado

Show 10 entries

Search

data_da_contratacao	cliente	cnpj	municipio	uf	subsetor_cnae_agrupado	setor_bndes	porte_do_cliente	valor_da_operacao_em_reais	cnpj_polui	cr
2016-01-04	NUTRIMAI S NUTRICAO E SAUDE LTDA - ME	21471658000151	ALFENAS	MG	Alojamento e alimentação	COMERCIO/SERVICOS	MICRO	10000.0	n	0.15
2016-01-12	RIO S I DO B C DE A CONGELADOS LTDA	15805459000130	RIO DE JANEIRO	RJ	Comércio	COMERCIO/SERVICOS	MÉDIA	7860.0	n	0.05
2016-01-12	DIONISIO OLIVEIRA CARDOSO E CIA LTDA	79799052000120	RONCADOR	PR	Comércio	COMERCIO/SERVICOS	MICRO	2500.1	n	0.25
2016-01-12	DIONISIO OLIVEIRA CARDOSO E CIA LTDA	79799052000120	RONCADOR	PR	Comércio	COMERCIO/SERVICOS	MICRO	8749.9	n	0.25
2016-01-13	ASSIS PASTEIS E SALGADOS LTDA - ME	15825378000100	MAUA	SP	Alojamento e alimentação	COMERCIO/SERVICOS	MICRO	8519.0	n	0.15
2016-01-13	ASSIS PASTEIS E SALGADOS LTDA - ME	15825378000100	MAUA	SP	Alojamento e alimentação	COMERCIO/SERVICOS	MICRO	1217.0	n	0.15
2016-01-19	DE-LA COMERCIO DE ALIMENTOS LTDA	17106502000102	BELO HORIZONTE	MG	Comércio	COMERCIO/SERVICOS	MICRO	5000.0	n	0.15
2016-01-26	ASSIS PASTEIS E SALGADOS LTDA - ME	15825378000100	MAUA	SP	Alojamento e alimentação	COMERCIO/SERVICOS	MICRO	800.0	n	0.15
2016-02-05	N.TUBOS LTDA ME	13151638000175	BLUMENAU	SC	Comércio	COMERCIO/SERVICOS	PEQUENA	9240.0	n	0.15
2016-02-15	J MOTTA E CIA LTDA	12573429000157	CASTANHAL	PA	Comércio	COMERCIO/SERVICOS	PEQUENA	9576.0	n	0.25

data\_da\_contratacao

cliente

cnpj

municipio

uf

subsetor\_cnae\_agrupado

setor\_bndes

porte\_do\_cliente

valor\_da\_operacao\_em\_reais

cnpj\_polui

cr



# Tecnologia e ferramentas utilizadas

---

- **Linguagem Python** para a extração e a organização dos dados, além da modelagem do Gaia;
- **Linguagem R** para a visualização;
- **Linguagem SQL** para a requisição de dados de CNPJ;
- **ShinyDashboard** para a criação do interface de visualização;
- **API de Google Maps Services** para a extração de geocode;
- **Leaflet** para a visualização de dados geolocalizados;
- **LightAutoML (LAMA)** para a construção do modelo **Gaia**;
- **Modelo BERT** para processar os textos em dados tabulares.

# Gratidão ao open-source!



HUGGING FACE

sberbank-ai-lab/  
**LightAutoML**

LAMA - automatic model creation framework



GAIA  
powered by  **BNDES**