



# Sistema de Previsão de Ocorrência de Descargas Elétricas Baseado em Aprendizado de Máquina

---

**Arielle Bassanelli**

Orientador: Prof.<sup>o</sup> Dr. Álvaro Fazenda  
Co-orientador: Prof.<sup>o</sup> Dr. Alan Calheiros

# Descargas Eléctricas no Brasil

Cerca de 70% dos desligamentos na transmissão de energia e 40% na distribuição são provocados por descargas eléctricas.



**Brasil, o país em que mais caem raios no mundo**

## Incidência de raios aumenta no Brasil no início de 2022

Só em janeiro, se formaram mais de 16 milhões de raios. De acordo com o Inpe, em 2021, 66 milhões de descargas eléctricas atingiram o Brasil, o país que tem a maior incidência de raios do mundo.

Por **Jornal Nacional**

04/02/2022 21h58 · Atualizado 04/02/2022



Sistema Eléctrico

# INTRODUÇÃO

## PREVISÃO E ALERTA DE EVENTOS SEVEROS



**Enxurrada**



**Alagamento**

- Região Sul e Sudeste frequentemente registram eventos severos.
- A severidade dos eventos severos pode estar relacionada a densidade de descargas elétricas observadas por satélite
- O Brasil é o país mais atingido por raios no mundo, cerca de 77,8 milhões de descargas elétricas por ano.
- 2012 foi o ano com maior incidência de raios no Brasil, registrando 94,3 milhões.

# OBJETIVO

Prever por meio de algoritmos de aprendizado de máquina regiões no Brasil associadas à atividade elétrica na atmosfera (raios).

Possível apoio à Órgãos do Governo na tomada de decisão para mitigação de efeitos e aprimoramento dos sistemas de alertas de eventos severos.

# OBJETIVOS ESPECÍFICOS

1. Combinar dados de modelos numéricos de previsão de tempo com dados observados de descargas elétricas
  - a. Base de dados para modelos de AM
2. Identificar as principais variáveis meteorológicas que caracterizam atividade elétrica;
3. Definir a severidade da atividade elétrica na atmosfera a partir dos dados do GOES-16
  - a. Densidade de ocorrência de descargas elétricas por  $\text{km}^2$ ;
4. Investigar o desempenho de técnicas de AM.

# QUESTÕES DE PESQUISA

- Quais são as principais variáveis atmosféricas que podem caracterizar a atividade elétrica atmosférica, e devem ser utilizadas no treinamento dos modelos de Inteligência Artificial?
- É possível a partir de técnica de Inteligência Artificial ou Aprendizado de Máquina, como por exemplo Rede Neural Artificial e Floresta Aleatória, identificar atividade elétrica na atmosfera sobre diferentes regiões do Brasil, a partir de dados oriundos de modelos numéricos de previsão de tempo?

# Formação de tempestades



# Nowcasting

- Definida por Keith Browning em 1981 como a descrição do estado atual do tempo e a previsão de mudanças esperadas em uma escala de tempo de algumas horas.
- Em 2010, a OMM limitou a previsão de tempo de *nowcasting* para um período de 6 horas.

## Informações Meteorológicas pretéritas e em tempo real!



**Radar Meteorológico de São Roque/SP - 31.01.2017 - 17:32h**

# Dados usuais para nowcasting

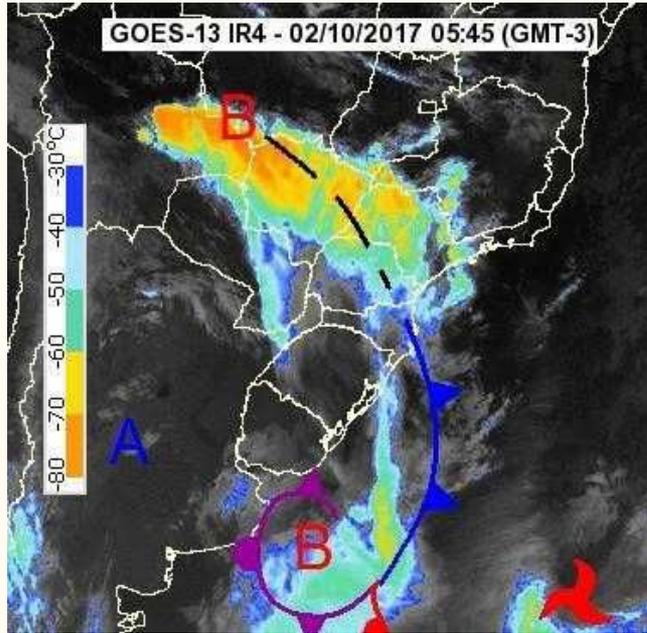
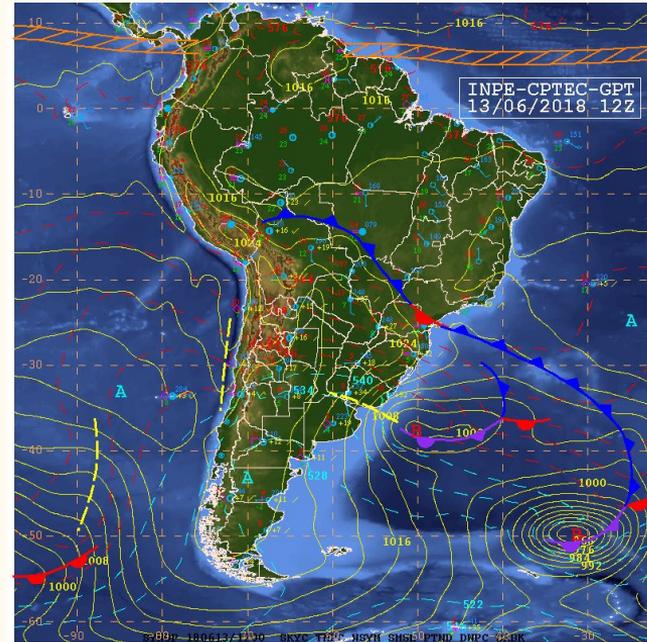


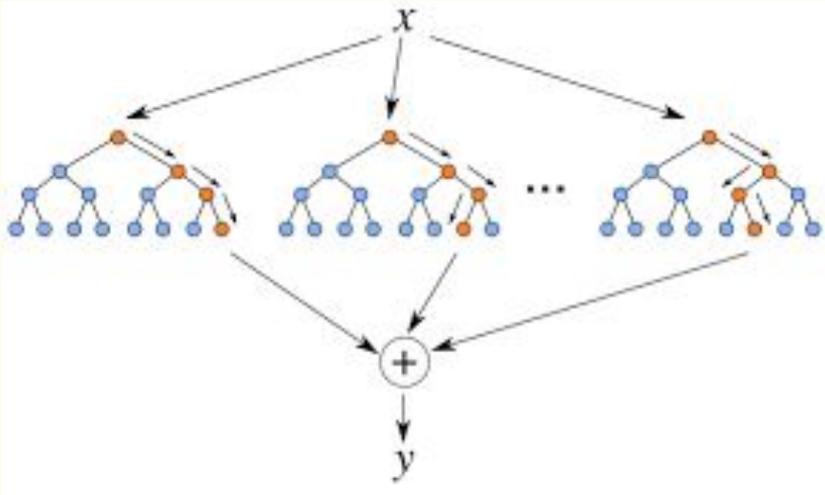
Imagem Satélite GOES-13  
02.10.2017 - 05:45h



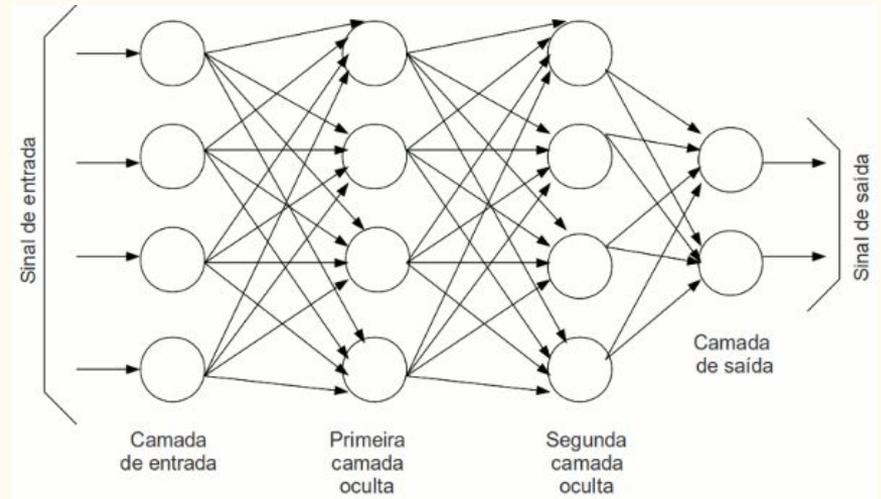
Modelo Numérico de Previsão de Tempo  
GFS 50km

# Métodos de aprendizado de máquina utilizados

## 1. Floresta Aleatória



## 2. Redes Neurais Artificiais



# Revisão Sistemática da Literatura

Base de Dados	Nº Total de Artigos	Nº de Artigos Selecionados
ACM	243	24
Science Direct	80	4
Scopus	1648	286
Adicionados	3	3
Duplicados	-12	-
<b>Total</b>	<b>1962</b>	<b>317</b>

Seleção dos Estudos

Base de Dados	Nº Total de Artigos	Nº de Artigos Selecionados
ACM	24	5
Science Direct	4	1
Scopus	286	21
Adicionados	3	3
<b>Total</b>	<b>317</b>	<b>30</b>

Extração dos Dados

- A previsão das descargas elétricas em alta resolução espacial e temporal continua sendo um desafio, pois poucos trabalhos tratam do assunto.
- O método de aprendizado de máquina mais utilizado para previsão de tempo são as redes neurais (60%);
  - As redes neurais, bem como muitos outros métodos, apresentam resultados satisfatórios apenas para previsão a curto prazo,
    - modelos probabilísticos que não consideram as condições físicas e dinâmicas da atmosfera com o passar do tempo;
  - O desempenho é dependente da qualidade das informações meteorológicas utilizadas em seu treinamento

# Base de Dados utilizada

Período entre 02/01/2019 e 24/01/2019 - América do Sul:

- WRF (CPTEC/INPE):
  - ◆ resolução espacial horizontal - 1 km;
  - ◆ previsão iniciada duas vezes ao dia: 00 e 12 UTC;
  
- GML: GOES-16 (CPTEC/INPE):
  - ◆ resolução espacial horizontal - 8x8 km;
  - ◆ uma imagem a cada 15 minutos.

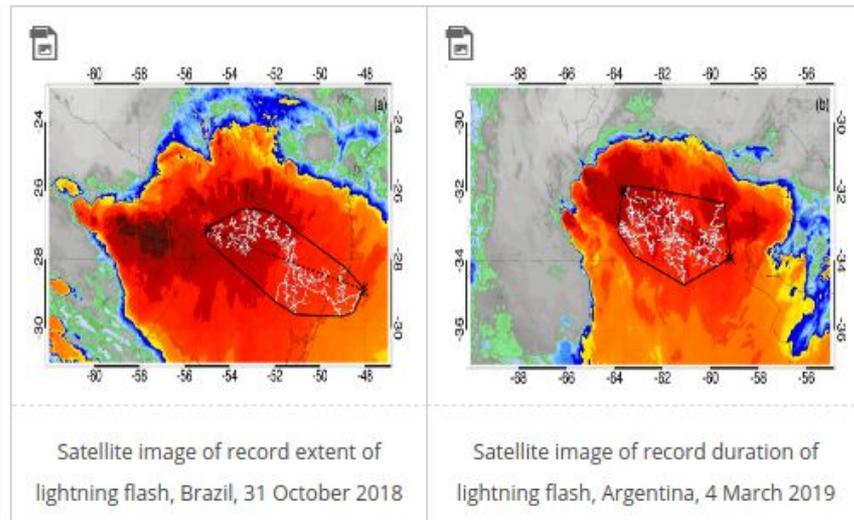
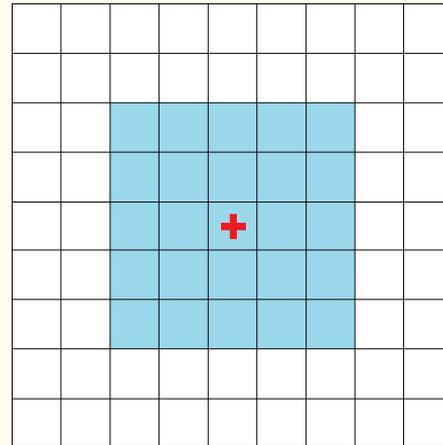


Imagem do GLM

# Extração de dados

Combinou-se dados de previsão do modelo WRF com dados observados do GLM, para os primeiros doze horários de cada simulação do WRF (00Z e 12Z), com os dados dos respectivos horários das observações e sua densidade de raios estimada (GLM).

Máscara com 2 pixels ao redor da coordenada com incidência de raio (grade GLM):



# Agrupamento (*clustering*)

- Método *K-Means* (*framework scikit-learn*)
- Agrupar dados em 3 diferentes classes de severidade de atividades elétricas
  - a partir da densidade de raios/m<sup>2</sup> do GLM
    - 1 - baixa atividade
    - 2 - média atividade
    - 3 - alta atividade

# Correlação para definição das variáveis do WRF

- Correlação entre todas as variáveis de saída do WRF com a densidade de raios/km<sup>2</sup> do GLM.
- Foram definidos quatro conjuntos:
  - Conjunto 1 - As primeiras 5 variáveis com o maior valor de correlação;
  - Conjunto 2 - As primeiras 10 variáveis com o maior valor de correlação;
  - Conjunto 3 - As primeiras 15 variáveis com o maior valor de correlação;
  - Conjunto 4 - Um conjunto de 12 variáveis escolhidos a partir de conhecimentos empíricos da área da meteorologia

# Aplicação dos métodos de ML

- Framework *sckit-learn*;
- Divisão de conjunto de treino e teste com “train\_test\_split”
  - 75% para treinamento e 25% para teste, sem embaralhamento dos dados;
- RNAs (binária e multiclasse) implementadas com “MLPClassifier”, com 3 camadas de 15 neurônios;
- Florestas Aleatórias com “RandomForestClassifier”, com profundidade máxima de 15;
- Avaliado o desempenho com os quatro diferentes conjuntos de variáveis do WRF

# Métricas de avaliação

1. FAR: representa a taxa de erro proveniente de falsos-positivos do classificador;
2. Precisão: mostra daqueles que foram classificados como corretos, quantos efetivamente eram;
3. POD: define a fração de eventos corretamente previstos;
4. *Bias*: representa o erro de aproximação de um modelo.

# RESULTADOS E DISCUSSÃO

1. Obteve-se 15.372 pontos de atividade com seus respectivos pontos de inatividade para análise (base balanceada), totalizando 30.743 amostras;
2. Dois tipos de classificação: Binária e Multi-classes;
3. Limiares das classes de severidade de atividade elétrica na atmosfera a partir do K-means:

<b>Categoria</b>	<b>Raios/km<sup>2</sup></b>
0 - sem atividade	0
1 - pouca atividade	0 >flash <= 17
2 - média atividade	18 >flash <= 66
3 - alta atividade	67 >flash

# RESULTADOS E DISCUSSÃO

4. Dentre as quinze melhores variáveis (campos meteorológicos) correlacionadas observa-se que muitas são variações do campo de vento, temperatura e umidade
  - a. De acordo com a RSL, as quatro variáveis mais utilizadas são: precipitação, vento, temperatura e umidade - 60% dos estudos
  - b. **Conjunto 2** (10 campos meteorológicos mais bem correlacionados) apresentou o melhor desempenho.
    - i. Hipótese: se mostra como um perfil da umidade até níveis médios na atmosfera, um ingrediente essencial para formação de nuvens em situações pré-convectivas (COTTON; BRYAN; C., 2010);
    - ii. conjunto 3 tem perfil similar - mais variáveis pode ter introduzido mais erros
    - iii. conjunto 1 limitou-se a regiões mais próximas da superfície, o que não deu representatividade às camadas onde a separação de cargas elétricas fica mais evidente;
    - iv. conjunto 4 apresentou o pior resultado

## 5. Resultados para os dias 09 a 11/09/2022

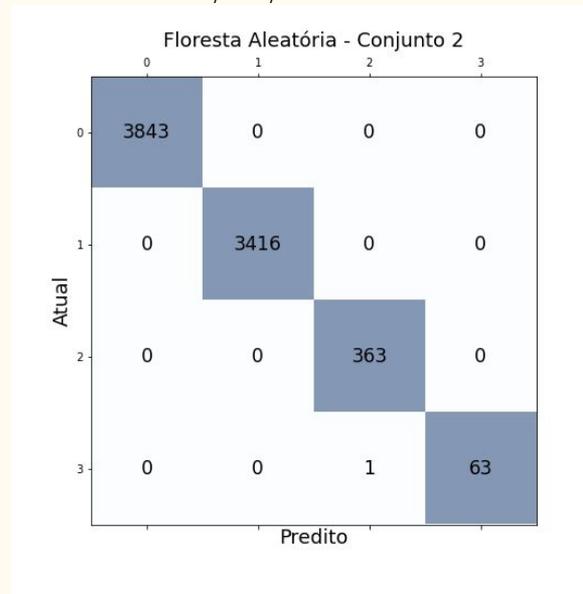
**Tabela 1. ML Models Results for the 4 sets.**

<b>ML Model</b>	<b>Set</b>	<b>FAR</b>	<b>Precision</b>	<b>POD</b>	<b>Bias</b>
Binary ANN	1	0,0000	1,0000	1,0000	1,0000
Binary ANN	2	0,0000	1,0000	1,0000	1,0000
Binary ANN	3	0,0010	0,9990	1,0000	1,0010
Binary ANN	4	0,7827	0,2173	0,6850	0,3153
MC ANN	1	0,0003	0,9997	1,0000	1,0003
MC ANN	2	0,0000	1,0000	1,0000	1,0000
MC ANN	3	0,0003	0,9997	0,9997	1,0000
MC ANN	4	0,0000	1,0000	0,5296	0,5296
RF	1	0,0000	1,0000	1,0000	1,0000
RF	2	0,0000	1,0000	1,0000	1,0000
RF	3	0,0000	1,0000	1,0000	1,0000
RF	4	0,0000	1,0000	1,0000	1,0000

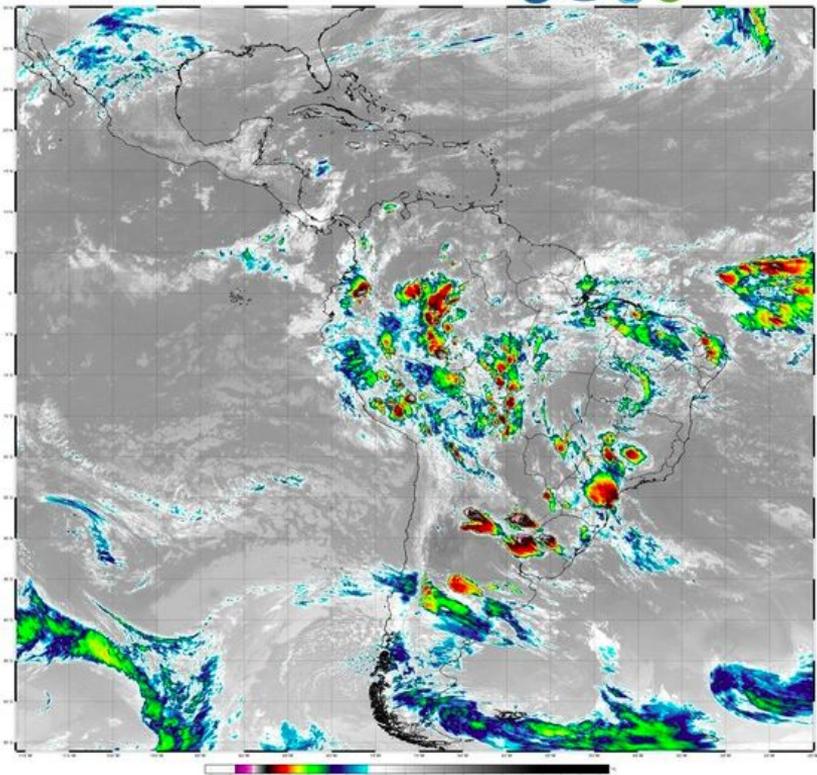
# RESULTADOS E DISCUSSÃO

6. O modelo de ML de melhor desempenho foi a **Floresta Aleatória**
  - a. diferenças sutis

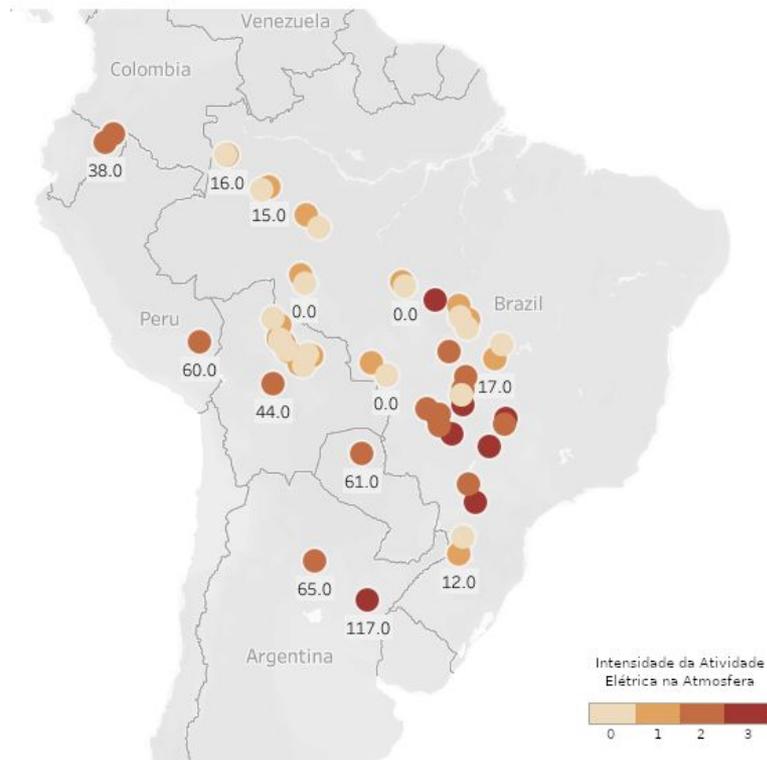
09/01/2019



GOES16 - CANAL 13 (10.30 microns)  
América Latina: 201901090000 - 201901090011 GMT



## Intensidade da Atividade Atmosférica em 24h - 20190109



# RESULTADOS E DISCUSSÃO

7. Teste com exclusão de dados fora do intervalo de confiança (75%) não teve sucesso
  - a. Exclui dados pouco frequentes de eventos mais severos foram excluídos

# CONSIDERAÇÕES FINAIS

- A RNA Binária treinada com os conjuntos 1 e 2 atingiu o objetivo de identificar as classes 0 e 1 com total êxito, acertando quase 100% dos casos do conjunto de teste;
- Dados podem estar relacionado com as condições atmosféricas pontuais e fortemente correlacionadas com a ocorrência de atividade elétrica,
  - Possível limitação da metodologia empregada;
- Modelos de AM apresentam resultados satisfatórios para previsão a curto prazo
  - Modelos probabilísticos
- A utilização de variáveis mais bem correlacionadas para construção de conjuntos de dados, se mostrou efetiva
  - não demandou conhecimentos específicos na área