



# ESTIMAÇÃO DA DISTÂNCIA DE OCORRÊNCIA DE RAIOS UTILIZANDO INTELIGÊNCIA COMPUTACIONAL E FORMAS DE ONDAS DE CAMPOS ELETROMAGNÉTICOS EMITIDOS POR DESCARGAS ATMOSFÉRICAS

Bruno Borges Guerra<sup>1</sup> – Unifesspa  
*brunoguerra@unifesspa.edu.br*  
Elton Rafael Alves<sup>2</sup> - Unifesspa  
*eltonrafaelalves@gmail.com*

**Agência Financiadora:** CNPq

**Eixo Temático/Área de Conhecimento:** Engenharia Elétrica / Engenharia IV

## 1. INTRODUÇÃO

A região amazônica no Brasil apresenta grande incidência de ocorrência de raios, principalmente a região sudeste do Pará, fazendo-se necessário a criação de sistemas de detecção e localização de raios. Assim, foi desenvolvido o sistema “*Lightning Detection and Waveform Storage System*” (LDWSS) que permite a ampliação do alcance dinâmico de detecção, possibilitando detectar e armazenar formas de onda de campos eletromagnéticos distantes e próximos emitidos por descargas atmosféricas. Além disso, tal sistema possui baixo custo quando comparado aos sistemas de monitoramento de descargas atmosféricas em LF (*Low Frequency*) em operação na região Amazônica. Porém, apesar dos grandes benefícios do sistema LDWSS, ainda não é possível identificar um padrão nas distâncias onde as descargas elétricas ocorreram, tornando-se limitado para realizar previsões, por exemplo, a ocorrência de um determinado raio pode ter sido a 50 km, mas também a 400 km.

Diante disso, justifica-se a importância do presente trabalho em aplicar inteligência computacional com o intuito de agrupar e posteriormente classificar as distâncias dos raios obtidas pelo LDWSS. Com isso, será possível identificar as distâncias de acordo com as formas de onda eletromagnética de cada raio.

## 2. MATERIAS E MÉTODOS

A metodologia deste trabalho consistiu inicialmente: obtenção e pré-processamento dos dados de formas de onda de raios através do LDWSS. Nesta etapa foi utilizadas técnicas normalização dos dados (min-max), bem como a utilização de filtro passa baixa para eliminação de frequências ruidosas.

Dentre os mais de oito mil eventos contidos no banco de dados, existem descargas nuvem-solo (negativos e positivos) e do tipo CID (*Compact Intracloud Discharge*). Tanto as formas de onda de raios nuvem-solo (CG) quanto dos CIDs foram observadas no período diurno e noturno. Dentre os raios nuvem-solo, encontram-se formas de onda de raios que ocorreram entre 6 e 500 km afastados do sensor. Esta pesquisa se concentrou nos raios nuvem-solo negativos.

A segunda etapa do projeto, consistiu no processo de clusterização das formas de onda, onde utilizou-se o Método K-means para identificar formas de ondas semelhantes, de acordo com a distância de ocorrência do raio. Na terceira etapa, realizou-se um processo de classificação dos grupos formados na etapa dois. Nessa etapa, realizou-se uma comparação entre quatro métodos de aprendizagem de máquina: K-nn, SVM (*Suporte*

---

<sup>1</sup>Graduando em Engenharia da Computação - Universidade Federal do Sul e Sudeste do Pará

<sup>2</sup>Doutor em Engenharia Elétrica – Professor Adjunto da Universidade Federal do Sul e Sudeste do Pará (FACEEL/IGE/Unifesspa).

# VI Seminário de Iniciação Científica

Pesquisa na Amazônia: Novos cenários

27 a 29 de Outubro de 2020

On-line pela plataforma Google Meet

UNIFESSPA | PROPIT

Vector Machine), Árvore de decisão e Naive Bayes. Essa comparação consistiu em avaliar qual método apresentaria maior acurácia de predição.

## 3. RESULTADOS E DISCUSSÃO

Através da linguagem Python utilizando do método de clusterização K-means, foi possível, a partir dos dados coletados no LWDSS, criar clusters contendo grupos de formas de onda dos raios relativas a uma certa distância. Foram criados 10 clusters para as formas de onda dos dados do Dia e 10 clusters para os dados da Noite, cada uma relativa uma distância média. E assim foi possível identificar e presumir uma forma de onda padrão para cada distância (linhas em vermelho nas Figura 1 e Figura 2).

Figura 1: Gráficos dos clusters dos dados do Dia

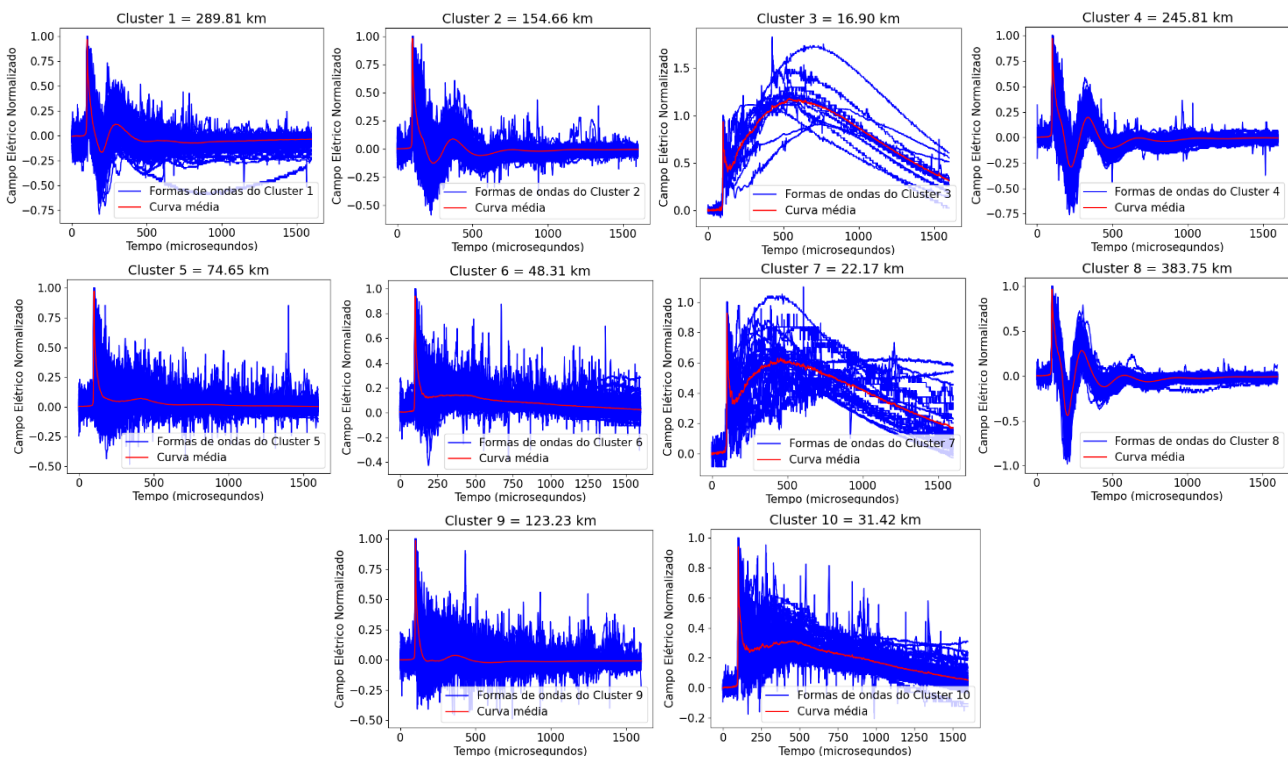
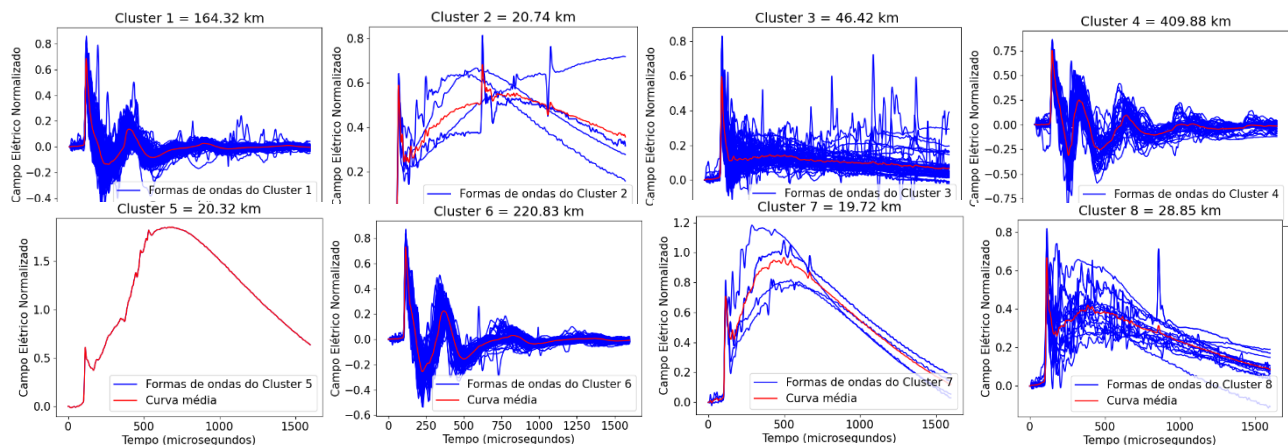


Figura 2: Gráficos dos clusters dos dados da Noite



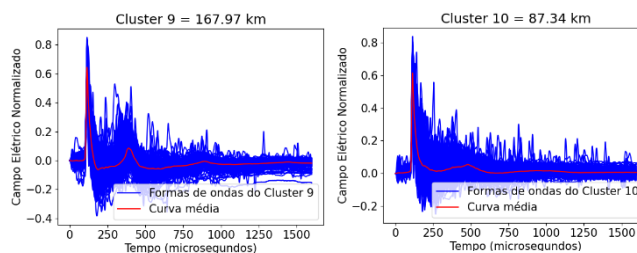
# VI Seminário de Iniciação Científica

Pesquisa na Amazônia: Novos cenários

27 a 29 de Outubro de 2020

On-line pela plataforma Google Meet

UNIFESSPA | PROPIT



A partir dos resultados obtidos na etapa de clusterização, utilizou-se métodos de Aprendizagem de Máquina para fazer-se a classificação da distância de ocorrência de raios, com base nos clusters obtidos para os dados de Dia e Noite. Utilizou-se quatro métodos de aprendizagem de máquina: K-nn, SVM, Árvore de decisão e Naive Bayes. Os dados foram divididos em 70% para treinamento e 30% para validação. A seguir será apresentado as matrizes de confusões (Tabelas 1 - 8) com as acurácias obtidas para cada cluster, considerando os métodos empregados.

- Método KNN

**Tabela 1: Matriz de confusão do Dia**

Real \ Predito	1	2	3	4	5	6	7	8	9	10	Todos
1	19	0	0	0	0	0	0	1	0	0	20
2	0	252	0	15	0	0	0	0	0	0	267
3	0	0	152	1	0	0	19	0	1	1	174
4	0	19	1	410	0	0	3	0	1	0	434
5	1	0	0	0	8	0	0	0	0	0	9
6	0	0	0	0	2	4	0	0	0	0	6
7	0	6	7	54	0	0	259	0	4	0	330
8	0	14	0	0	0	0	0	77	0	0	91
9	0	0	0	27	0	0	2	0	110	1	140
10	0	0	17	0	0	0	0	0	8	77	102
Todos	20	291	177	507	10	4	283	78	124	79	1573

► Acurácia: 87%

**Tabela 2: Matriz de Confusão da Noite**

Real \ Predito	1	2	3	4	5	6	7	8	9	10	Todos
1	23	0	0	0	0	3	0	0	0	0	26
2	0	4	0	0	0	0	0	0	0	0	4
3	0	0	107	0	0	0	1	0	0	0	108
4	0	0	0	0	0	0	0	0	0	0	0
5	2	0	1	0	8	0	0	0	0	0	11
6	2	0	4	0	1	67	1	0	0	0	75
7	0	0	5	0	0	5	52	0	0	0	62
8	0	0	0	1	0	0	0	0	0	0	1
9	0	0	2	0	0	0	0	0	17	0	19
10	0	2	0	0	0	0	0	0	0	0	2
Todos	27	6	119	1	9	75	54	0	17	0	308

► Acurácia: 90%

- Método SVM

**Tabela 3: Matriz de confusão do Dia**

Real \ Predito	1	2	3	4	5	6	7	8	9	10	Todos
1	18	0	0	0	1	0	0	1	0	0	20
2	0	256	0	10	0	0	1	0	0	0	267
3	0	0	164	0	0	0	9	0	0	1	174
4	0	2	0	423	0	0	9	0	0	0	434
5	0	0	0	0	9	0	0	0	0	0	9
6	0	0	0	0	3	3	0	0	0	0	6
7	0	2	3	2	0	0	323	0	0	0	330
8	0	6	0	0	0	0	0	85	0	0	91
9	0	0	0	2	1	0	0	0	137	0	140
10	0	0	5	0	0	0	0	0	5	92	102
Todos	18	266	172	437	14	3	342	86	142	93	1573

► Acurácia: 96%

**Tabela 4: Matriz de confusão da Noite**

Real \ Predito	1	2	3	4	5	6	7	8	9	10	Todos
1	23	0	0	0	0	3	0	0	0	0	26
2	0	4	0	0	0	0	0	0	0	0	4
3	0	0	108	0	0	0	0	0	0	0	108
4	0	0	0	0	0	0	0	0	0	0	0
5	2	0	0	0	9	0	0	0	0	0	11
6	0	0	3	0	1	71	0	0	0	0	75
7	0	0	1	0	0	5	56	0	0	0	62
8	0	0	0	0	1	0	0	0	0	0	1
9	0	1	2	0	0	0	0	0	16	0	19
10	0	0	0	0	0	0	0	0	0	2	2
Todos	25	5	114	0	11	79	56	0	16	2	308

► Acurácia: 94%

- **Método Arvore de Decisão**

**Tabela 5: Matriz de confusão do Dia**

Real \ Predito	1	2	3	4	5	6	7	8	9	10	Todos
1	17	1	0	0	1	0	0	1	0	0	20
2	0	197	0	49	0	0	12	9	0	0	267
3	0	0	129	9	0	0	21	0	3	12	174
4	0	44	2	316	0	0	45	0	25	2	434
5	0	0	0	0	9	0	0	0	0	0	9
6	0	0	0	0	1	5	0	0	0	0	6
7	0	8	21	51	0	0	240	0	9	1	330
8	2	10	0	0	0	0	0	79	0	0	91
9	0	4	7	21	0	0	14	0	80	14	140
10	0	0	8	3	0	0	1	0	13	77	102
Todos	19	264	167	449	11	5	333	89	130	106	1573

► Acurácia: 73%

**Tabela 6: Matriz de confusão da Noite**

Real \ Predito	1	2	3	4	5	6	7	8	9	10	Todos
1	19	0	1	0	0	4	2	0	0	0	26
2	0	4	0	0	0	0	0	0	0	0	4
3	2	0	92	0	0	6	3	0	5	0	108
4	0	0	0	0	0	0	0	0	0	0	0
5	1	0	3	0	4	0	3	0	0	0	11
6	4	0	7	0	0	55	9	0	0	0	75
7	1	0	2	0	0	6	53	0	0	0	62
8	0	0	0	1	0	0	0	0	0	0	1
9	0	1	1	0	0	0	0	0	17	0	19
10	0	0	0	0	0	0	0	0	0	2	2
Todos	27	5	106	1	4	71	70	0	22	2	308

► Acurácia: 80%

- **Método Naive Bayes**

**Tabela 7: Matriz de Confusão do Dia**

Real \ Predito	1	2	3	4	5	6	7	8	9	10	Todos
1	19	0	0	0	1	0	0	0	0	0	20
2	0	221	0	25	0	0	7	9	5	0	267
3	0	1	138	0	0	0	12	0	14	9	174
4	0	35	3	342	0	0	28	0	26	0	434
5	0	0	0	0	9	0	0	0	0	0	9
6	0	0	0	0	3	3	0	0	0	0	6
7	0	11	22	25	0	0	247	0	25	0	330
8	3	10	0	0	0	0	0	78	0	0	91
9	0	0	4	6	1	0	6	0	98	25	140
10	0	0	7	0	0	0	0	0	5	90	102
Todos	22	278	174	398	14	3	300	87	173	124	1573

► Acurácia: 79%

**Tabela 8: Matriz de confusão da Noite**

Real \ Predito	1	2	3	4	5	6	7	8	9	10	Todos
1	21	0	0	0	0	4	1	0	0	0	26
2	0	4	0	0	0	0	0	0	0	0	4
3	0	0	94	0	0	4	4	0	6	0	108
4	0	0	0	0	0	0	0	0	0	0	0
5	1	0	0	0	10	0	0	0	0	0	11
6	2	0	5	0	1	57	10	0	0	0	75
7	2	0	5	0	6	15	34	0	0	0	62
8	0	0	0	1	0	0	0	0	0	0	1
9	0	2	1	0	0	0	0	0	16	0	19
10	0	0	0	0	0	0	0	0	0	2	2
Todos	26	6	105	1	17	80	49	0	22	2	308

► Acurácia: 77%

Pelos resultados obtidos, observou-se que a melhor predição obtida foi através do emprego do método SVM, com uma acurácia de 94% para dados da Noite e 96% para dados do Dia.

#### 4. CONSIDERAÇÕES FINAIS

O LDWSS, possibilita da captação de dados das descargas atmosféricas na microrregião do sul e sudeste do Pará, essas informações e os conhecimentos obtidos durante o desenvolvimento do projeto permitiu resultados a respeito separação dos dados tornando-os informações de fácil leitura, permitindo extrair dados estatísticos, como a quantidade de eventos diários e categoria dos raios a partir de uma visualização mais detalhada do gráfico. Dados esses que possibilitaram a criação de categorias de raios para uma determinada distância e a partir disso, permitir, através de métodos de aprendizagem de máquina, que novos dados possam ser categorizados automaticamente.

#### REFERÊNCIAS

A. F. R. Leal, V. A. Rakov, and B. R. P. da Rocha, "Upgrading the Low-Cost Lightning Detection and Waveform Storage System," *IEEE Trans. Electromagn. Compat.*, 2018.