

# Instituto Senai de Inovação em Sistemas Embarcados

## SELETIVO PARA PESQUISADOR DE I.A

### Definição do problema

Linhas de tensão percorrem centenas de quilômetros para fornecer energia às cidades. Essas grandes distâncias tornam o trabalho de inspecionar manualmente as linhas em busca de danos uma tarefa complexa e cara. Esses danos levam a um fenômeno conhecido como descarga parcial - descargas parciais são descargas elétricas que ocorrem em defeitos no isolamento de cabos, também podem ocorrer nas suas interfaces (isolante e semicondutora), ou em emendas e terminais. Entre suas causas estão o envelhecimento dos componentes, contaminação do material isolante, falha de instalação, entre outras. As descargas parciais danificam gradativamente as linhas de tensão, portanto, se não forem reparadas, podem levar a uma queda de energia, danificar equipamentos ou provocar desastres ambientais.

Você foi contratado para desenvolver um modelo de Aprendizado de Máquina para identificar padrões nos sinais de Linhas elétricas e com isso, detectar a ocorrência ou não de descargas parciais. Desenvolva o Pipeline completo de um projeto de Aprendizado de Máquina, desde a Análise e preparação dos dados adquiridos até a disponibilização do algoritmo em Container (utilizando Docker) no GitHub.

O candidato é livre para usar a criatividade sobre qual caminho seguir e como utilizar os dados para a resolução do problema.

### Descrição dos dados

A base de dados contém 501 medições de sinais

([https://drive.google.com/file/d/17LPP\\_ZZprHIG3R2f0Vhg5CyfFZOMvlia/view?usp=sharing](https://drive.google.com/file/d/17LPP_ZZprHIG3R2f0Vhg5CyfFZOMvlia/view?usp=sharing)), cada sinal contém 800.000 medições da tensão de uma linha de energia (representado pelas colunas). A própria rede opera em um esquema de energia trifásico (0,1,2) e todas as três fases são medidas simultaneamente. As últimas 3 colunas da base de dados representam respectivamente o ID do sinal, fase, e o target representando ausência (0) ou presença de descarga parcial (1).

	0	1	2	3	4	5	6	7	8	9	...	799993	799994	799995	799996	799997	799998	799999	signal_id	phase	target
0	18	18	17	18	18	18	19	18	18	17	...	17	18	19	19	17	19	17	0	0	0
1	1	0	-1	1	0	0	1	0	0	0	...	0	0	2	1	0	1	0	1	1	0
2	-19	-19	-20	-19	-19	-20	-18	-19	-20	-19	...	-18	-19	-18	-19	-19	-18	-19	2	2	0
3	-16	-17	-17	-16	-16	-15	-16	-17	-18	-17	...	-15	-15	-15	-15	-15	-14	-14	3	0	1
4	-5	-6	-6	-5	-5	-4	-5	-7	-7	-7	...	-5	-4	-4	-4	-4	-3	-4	4	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
495	-26	-21	-24	-25	-23	-26	-22	-23	-25	-20	...	-22	-20	-21	-20	-22	-22	-20	495	0	0
496	9	13	8	10	10	8	12	10	11	13	...	12	12	13	13	13	13	11	496	1	0
497	8	11	7	9	8	7	10	9	9	10	...	9	10	10	11	10	10	10	497	2	0
498	-18	-18	-19	-19	-19	-18	-17	-18	-19	-18	...	-19	-19	-19	-19	-19	-20	-19	498	0	0
499	1	0	-1	-1	0	0	0	-1	-1	0	...	0	0	0	0	0	-1	0	499	1	0

# Avaliação

A prova prática terá duas etapas: a **entrega do projeto** e uma **entrevista** para a apresentação do projeto.

Itens avaliados na entrega do projeto:

- Organização do código
- Estruturação do projeto no GitHub
- Interpretação do problema
- Etapas do Pipeline realizadas e qualidade de sua execução
  - *Pré-processamento, Engenharia de Requisitos, Seleção do Modelo, Avaliação do Modelo, disponibilização em Container Docker.*

Itens avaliados na entrevista

- Apresentação do projeto e defesa do ponto de vista.
- Visualização dos dados e resultados

**OBS1:** O candidato deve submeter seu projeto independentemente da não realização de uma das etapas do exigidas.

**OBS2:** A apresentação do projeto pode ser feita utilizando qualquer ferramenta (power point, jupyter notebook, Google Colab etc.)

## Requisitos de Software

É permitido o uso de qualquer IDE e Biblioteca de Aprendizado de Máquina, desde que respeite às seguintes exigências:

- Linguagem de Programação Python, R ou C++
- Docker

## Especificações de Hardware

Recomendado usar um computador com:

- 8GB de memória RAM ou superior
- Processador Core I5 ou superior